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

Ying Xia[a] Peter C. B. Phillips[b] Liangjun Su[c]

[a]School of Economics, Singapore Management University, Singapore
[b]Yale University, University of Auckland, Southampton University and Singapore Management University
[c]School of Economics and Management, Tsinghua University, China

November 12, 2022

Abstract

In this article, we consider a generalized panel data transformation model with fixed effects where the structural functions are assumed to be additive. Our model does not impose parametric assumptions on the transformation function, the structural function, or the distribution of the idiosyncratic error term. We propose a multiple-stage Local Maximum Likelihood Estimator (LMLE) for the structural functions. 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 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.

Key Words: Logit sieve estimation, matching, nonparametrics, oracle efficiency, panel data, structural function, transformation model.

JEL classification: C13, C14, C23

*Phillips acknowledges support from a Lee Kong Chian Fellowship at SMU, the Kelly Fund at the University of Auckland, and the NSF under Grant No. SES 18-50860. Su gratefully acknowledges the support from the National Natural Science Foundation of China under Grant No. 72133002. Address correspondence to: Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, 100084, China. Email: sulj@sem.tsinghua.edu.cn.

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, Komunjer and Kristensen (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, Pan, Lv and Zhang (2018), among others. Recently, Chen, Lu and Wang (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.

2 Methodology

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

2.1 The Model

We consider the following transformation model:

$$\Lambda(Y_{it}) = g(X_{it}) + \alpha_i + \epsilon_{it} = \sum_{l=1}^{d} g_l(X_{l,it}) + \alpha_i + \epsilon_{it}, \tag{2.1}$$

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, $g(X_{it}) = \sum_{l=1}^{d} g_l(X_{l,it})$, α_i is the individual fixed effect that captures the unobserved individual heterogeneity, ϵ_{it} is the idiosyncratic error term, and $\Lambda(\cdot)$ is an unknown transformation function that is strictly increasing. Note that the model in (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, $\dot{g}_l(X_{l,it})$ can be interpreted as the marginal effect of $X_{l,it}$ on $\Lambda(Y_{it})$. 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 (2.1) as follows:

$$Y_{it} = \Lambda^{-1}(g(X_{it}) + \alpha_i + \epsilon_{it}) = \Lambda^{-1}\left(\sum_{l=1}^{d} g_l(X_{l,it}) + \alpha_i + \epsilon_{it}\right), \tag{2.2}$$

where $\Lambda^{-1}(\cdot)$ is the inverse function of $\Lambda(\cdot)$. Clearly, the above expression indicates that the model (2.1) is different from the classical panel data model of the following form:

$$Y_{it} = \Lambda^{-1}(g(X_{it})) + \alpha_i + \epsilon_{it} = \Lambda^{-1}\left(\sum_{l=1}^{d} g_l(X_{l,it})\right) + \alpha_i + \epsilon_{it}. \tag{2.3}$$

For the model in (2.3), we can eliminate the fixed effects through various transformations such as first-differing and within-group transformation. Nevertheless, for the model in (2.1) or (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 $D_{i,ts} \equiv \mathbf{1}\{Y_{it} > Y_{is}\}$ is free of α_i. This motivates the estimation of the structural functions based on such a non-smooth transformation of the dependent variables.

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 $D_i \equiv \mathbf{1}\{Y_{i2} > Y_{i1}\}$.

Since $\Lambda(\cdot)$ is strictly increasing, we have

$$\begin{aligned} D_i &= \mathbf{1}\left\{\Lambda\left(Y_{i2}\right) > \Lambda\left(Y_{i1}\right)\right\} \\ &= \mathbf{1}\left\{g\left(X_{i2}\right) + \epsilon_{i2} > g\left(X_{i1}\right) + \epsilon_{i1}\right\} \\ &= \mathbf{1}\left\{g\left(X_{i2}\right) - g\left(X_{i1}\right) > \Delta_i\right\} \\ &= \mathbf{1}\left\{\sum_{l=1}^{d} g_l\left(X_{l,i2}\right) - \sum_{l=1}^{d} g_l\left(X_{l,i1}\right) > \Delta_i\right\}, \end{aligned} \tag{2.4}$$

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

$$E\left(D_i|X_i\right) = \Pr\left(g\left(X_{i2}\right) - g\left(X_{i1}\right) > \Delta_i\right) = F\left(g\left(X_{i2}\right) - g\left(X_{i1}\right)\right). \tag{2.5}$$

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.

2.2.1 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 (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 (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:

$$\begin{aligned}\Delta g(X_i) &= \left(P^K(X_{i2}) - P^K(X_{i1})\right)'\beta_0 + [r_1(X_{i2}) - r_1(X_{i1})] \\ &\equiv \Delta P^K(X_i)'\beta_0 + \Delta r_1(X_i) \qquad (2.6)\end{aligned}$$

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

$$\begin{aligned}LF(\Delta g(X_i)) &= LF\left(\Delta P^K(X_i)'\beta_0 + \Delta r_1(X_i)\right) \\ &= LF\left(\Delta P^K(X_i)'\beta_0\right) + LF(\Delta g_i^*)\Delta r_1(X_i) \\ &= \sum_{\ell=0}^R \alpha_{\ell,0}\left(\Delta P^K(X_i)'\beta_0\right)^\ell + [r_2(X_i) + LF(\Delta g_i^*)\Delta r_1(X_i)] \\ &\equiv \sum_{\ell=0}^R \alpha_{\ell,0}\left(\Delta P^K(X_i)'\beta_0\right)^\ell + r(X_i), \qquad (2.7)\end{aligned}$$

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 $LF(\Delta g(X_i))$:

$$1,\ \Delta P^K(X_i),\ \text{the tensor product of } \Delta P^K(X_i) \text{ up to order } R. \qquad (2.8)$$

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:

$$\bar{\pi} = \arg\min_{\pi} -\frac{1}{n}\sum_{i=1}^{N}\left[D_i \cdot \ln\left(L\left(R(X_i)'\pi\right)\right) + (1-D_i)\cdot \ln\left(1 - L\left(R(X_i)'\pi\right)\right)\right] + \lambda \left\|\pi\right\|_1, \tag{2.9}$$

where $L(\cdot)$ is the Logit CDF: $L(x) = \dfrac{\exp(x)}{1+\exp(x)}$, $\|\cdot\|_1$ 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

$$\lambda = cn^{-1/2}\Phi^{-1}\left(1 - c_{\lambda n}/\left\{2K_R\right\}\right) \tag{2.10}$$

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 $\bar{m}_i \equiv \bar{E}(D_i|X_i) = L(R(X_i)'\bar{\pi})$, which serves as an initial consistent estimator for $m_i \equiv E(D_i|X_i)$. 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 (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

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

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

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

So in principle, one can consider minimizing the average squared distance between $\Delta g(X_i)$ and $\Delta g(X_j)$ when we control m_i to lie close to m_j. In practice, both $\Delta g(X_i)$ 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

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

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 $\beta^0 = \left(1, \theta^{0\prime}\right)'$. The matching estimator of θ^0 is obtained as follows:

$$\begin{aligned} \bar{\theta} &= \arg\min_{\theta} \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \left[\triangle P_{i,j}^{1,K} + \theta' \triangle P_{i,j}^{K-1,K}\right]^2 H_{1h_1}\left(\bar{m}_j - \bar{m}_i\right) & (2.11) \\ &= -\left\{\frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1,ji}\right\}^{-1} \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \triangle P_{i,j}^{1,K} \triangle P_{i,j}^{K-1,K} \bar{H}_{1h_1,ji} & (2.12) \end{aligned}$$

where $\bar{H}_{1h_1,ji} = H_{1h_1}\left(\bar{m}_j - \bar{m}_i\right)$, $H_{1h_1}(\cdot) \equiv h_1^{-1} H_1\left(\cdot/h_1\right)$, $H_1(\cdot)$ 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 $g_l\left(x_l\right)$ by $\bar{g}_l(x_l) = p^K(x_l)' \bar{\beta}^{x_l}$ for $l = 1, \cdots, d$ and that of $g\left(x\right)$ by $\bar{g}(x) \equiv \sum_{l=1}^d \bar{g}_l(x_l)$, where $x = \left(x_1, ..., x_d\right)'$.

2.2.2 Second-stage estimation

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

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

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

$$\begin{aligned} & \sum_{i=1}^{N} \left\{D_i \ln\left[F\left(\Delta g_i\right)\right] + \left(1 - D_i\right) \ln\left[1 - F\left(\Delta g_i\right)\right]\right\} \\ &= \sum_{i=1}^{N} \left\{D_i \ln\left[L\left(LF\left(\Delta g_i\right)\right)\right] + \left(1 - D_i\right) \ln\left[1 - L\left(LF\left(\Delta g_i\right)\right)\right]\right\}. & (2.13) \end{aligned}$$

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

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

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

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

$$\begin{aligned} & Q_n\left(\triangle \bar{g}_j, \{b_l\}_{l=0}^{a_2}\right) \\ &= \frac{-1}{n} \sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right) \end{aligned}$$

$$\times\left\{D_i \ln\left[L\left(b_0+\sum_{l=1}^{a_2}\frac{1}{h_2^l l!}\left(\triangle\bar{g}_{i,j}\right)^l b_l\right)\right]+(1-D_i)\ln\left[1-L\left(b_0+\sum_{l=1}^{a_2}\frac{1}{h_2^l l!}\left(\triangle\bar{g}_{i,j}\right)^l b_l\right)\right]\right\}.$$

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, $b_0+\sum_{l=1}^{a_2}\frac{1}{h_2^l l!}\left(\triangle\bar{g}_{i,j}\right)^l b_l$ serves as an a_2-order Taylor series approximation of $LF\left(\Delta\bar{g}_i\right)$ in the neighborhood of $\triangle\bar{g}_j$. Then we can estimate $\left(LF_j, h_2\dot{LF}_j\right)$ by the minimizing $Q_n\left(\triangle\bar{g}_j,\{b_l\}_{l=0}^{a_2}\right)$ 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 $\widehat{\dot{LF}}_j=\widehat{\partial LF}_j$.

2.2.3 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 $\Delta g_i=\sum_{j=1}^{d}\left[g_j\left(X_{j,i2}\right)-g_j\left(X_{j,i1}\right)\right]$ 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 $g_l\left(X_{l,i2}\right)$ and $g_l\left(X_{l,i1}\right)$ 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 $g_l\left(X_{l,i2}\right)$ and $g_l\left(X_{l,i1}\right)$ separately around a point x_l below.

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

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

and

$$LF\left(G_{l1,i}\right)\approx LF(\Delta g_i)-\dot{LF}(\Delta g_i)\left[g_l\left(x_l\right)+\dot{g}_l\left(x_l\right)\left(X_{l,i1}-x_l\right)-g_l\left(X_{l,i1}\right)\right]\equiv LF_{i1}(x_l).\qquad(2.14)$$

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

$$\sum_{j=1,j\neq l}^{d}\left[g_j\left(X_{j,i2}\right)-g_j\left(X_{j,i1}\right)\right]+g_l\left(x_l\right)+\dot{g}_l\left(x_l\right)\left(X_{l,i2}-x_l\right)-g_l\left(X_{l,i1}\right)$$

$$= \Delta g_i + g_l(x_l) + \dot{g}_l(x_l)(X_{l,i2} - x_l) - g_l(X_{l,i2}) \equiv G_{l2,i},$$

and

$$LF(G_{l2,i}) \approx LF(\Delta g_i) + \dot{LF}(\Delta g_i)\left[g_l(x_l) + \dot{g}_l(x_l)(X_{l,i2} - x_l) - g_l(X_{l,i2})\right] \equiv LF_{i2}(x_l). \quad (2.15)$$

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 $(g_l(x_l), \dot{g}_l(x_l))$

$$\begin{aligned} &\sum_{t=1}^{2}\sum_{i=1}^{N} H_{3h_3}(X_{l,it} - x_l)\left\{D_i \ln\left[L(LF(G_{lt,i}))\right] + (1 - D_i)\ln\left[1 - L(LF(G_{lt,i}))\right]\right\} \\ &\approx \sum_{t=1}^{2}\sum_{i=1}^{N} T_{h_3}(X_{l,it} - x_l)\left\{D_i \ln\left[L(LF_{it}(x_l))\right] + (1 - D_i)\ln\left[1 - L(LF_{it}(x_l))\right]\right\}, \end{aligned} \quad (2.16)$$

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 (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

$$W_{n,x_l}(c) \equiv -\sum_{t=1}^{T}\sum_{i=1}^{N} H_{3h_3}(X_{l,it} - x_l)\left[D_i \ln\left(L\left(\widehat{LF}_{it,x_l}(c)\right)\right) + (1 - D_i)\ln\left(1 - L\left(\widehat{LF}_{it,x_l}(c)\right)\right)\right], \quad (2.17)$$

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

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

By minimizing the objective function in (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 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.

3.1 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 1 *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 $\underline{c}_\epsilon \leq E\left(\epsilon_{it}^2\right) \leq \overline{c}_\epsilon$ and $E|\epsilon_{it}|^j \leq c_\epsilon^{j-2} j! E\left(\epsilon_{it}^2\right) < \infty$ for all $j \geq 2$.

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

Assumption 2 *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 3 *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* (Π) *for some compact set* Π *in* $\mathbb{R}^{K_R}$ *such that* $\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}} |LF(\Delta g(x)) - R(x)'\pi^0| = O(K^{-\gamma} + R^{-(R+1)})$; *we can decompose* $R(x) = (R_1(x)', R(x)')'$ *and* $\pi^0 = \left(\pi_1^{0\prime}, \pi_2^{0\prime}\right)'$ *accordingly such that* $s_{\pi_1}^2 \log^2\left(K^R \vee n\right) \leq K^{-\gamma} n$, *and* $\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}} \left|R_2(x)'\pi^0\right| = O(K^{-\gamma})$ *where* $s_{\pi_1} \equiv \left|\pi_1^0\right|_0$.

Assumption 3(1) imposes some smooth conditions on $F(\cdot)$ to ensure the second sieve approximation considered in the first stage estimation. Assumption 3(2) ensures the desirable asymptotic properties of the sieve logit estimator in the first stage. Assumption 3(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 3(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 3(5) indicates that π^0 should be approximately sparse to facilitate the asymptotic analysis.

Assumption 4 *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* $\eta(m_i) \equiv E\left[\triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime}|m_i\right]$ *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:* $0 < C_{1L} \leq \lambda_{\min}(E[\eta(m_i) f_m(m_i)]) \leq \lambda_{\max}(E[\eta(m_i) f_m(m_i)]) \leq C_{2L} < \infty$.

Assumption 4(1) impose some standard conditions to ensure the logit sieve estimator to be well behaved. Assumption 4(2) ensures the matching estimator in the second substep of the first stage estimation is well behaved.

Assumption 5 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 5 imposes some standard conditions on the density of the regressors.

Assumption 6 *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 6(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 6(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 7 *1. As $n \to \infty$, $K \to \infty$, $R \to \infty$, $h_\ell \to 0$ $\forall \ell \in [3]$, and $R^{-(R+1)} = O(K^{-\gamma})$;*

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

Assumption 7 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 7(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 7(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 $g_l(\cdot)$ 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 7(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 $K = n^{1/5}$.

3.2 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 3.1 *Suppose that Assumptions 1–4, 6(1) and 7(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)

$$\bar{\theta} - \theta_0 = -\left\{ \frac{1}{n^2} \sum_{1 \le i \ne j \le n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1,ji} \right\}^{-1} \left\{ \frac{1}{n^2} \sum_{1 \le i \ne j \le n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} H_{1h_1,ji} \right.$$

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

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

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

Theorem 3.1(i) establishes a Bahadur-type representation for the first-stage estimator $\bar{\theta}$. Theorem 3.1(ii) establishes the Euclidean norm for $\bar{\theta}$. Theorem 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 3.2 *Suppose that Assumptions 1–4, 6 and 7(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*

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

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 3.2(i) establishes the asymptotic expansions for $\widehat{LF}(\triangle g_j))$ and $h_2 \widehat{\partial LF}(\triangle g_j)$; Theorem 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 3.1 and 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 *Suppose that Assumptions 1–7 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)

$$\begin{pmatrix} \sqrt{nh_3} & 0 \\ 0 & \sqrt{nh_3^3} \end{pmatrix} \left(\begin{pmatrix} \widehat{g}_l(x_l) \\ \widehat{\dot{g}}_l(x_l) \end{pmatrix} - \begin{pmatrix} g_l(x_l) \\ \dot{g}_l(x_l) \end{pmatrix} \right) - \frac{1}{2} \ddot{g}_l(x_l) \begin{pmatrix} h_3^2 \kappa_{21} \\ 0 \end{pmatrix}$$

$$\xrightarrow{d} N\left(0, \frac{1}{2}E\left\{\left.\frac{F(\triangle g(X_i))\left[1-F(\triangle g(X_i))\right]}{\dot{F}^2(\triangle g(X_i))}\right| X_{l,it}=x_l\right\}\begin{pmatrix}\kappa_{02} & 0\\ 0 & \frac{\kappa_{22}}{\kappa_{21}^2}\end{pmatrix}\right);$$

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

Theorem 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(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(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, 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, Chetverikov and Kato (2014, Corollary 3.1). For brevity, we skip the details.

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. $\alpha_i = 0.5(X_{1,it}+X_{2,it})+0.5\eta$, 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 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

Table 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

Table 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

Table 1 - 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 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.

Appendices

This appendix is composed of two sections. Section A contains the proofs of the main results in the paper. Section B contains some technical lemmas that are used in the proof of the main results and their proofs.

A Proofs of the Main Results

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

A.1 Proof of Theorem 3.1

A.1.1 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 (2.7) and the definition of $R\left(\cdot\right)$, we have

$$\begin{aligned} E\left(D_i|X_i\right) &= F\left(\Delta g\left(X_i\right)\right)=L(LF(\Delta g\left(X_i\right)) \\ &= L\left(R(X_i)'\pi^0+r(X_i)\right)=L\left(R(X_i)'\pi^0\right)+\dot{L}\left(R_i^*\right)r(X_i) \\ &\equiv L\left(R(X_i)'\pi^0\right)+r_L(X_i), \end{aligned} \tag{A.1}$$

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 $\sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}}|r(x)|=O\left(K^{-\gamma}\right)$ under Assumptions 3 and 7.

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

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

Under Assumption 4(1), we can show that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\left[R(X_i)'\left(\bar{\pi}-\pi^0\right)\right]^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\bar{\pi}-\pi^0\right)'R(X_i)R(X_i)'\left(\bar{\pi}-\pi^0\right) \\ &\geq \left(\bar{\pi}-\pi^0\right)'\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}R(X_i)R(X_i)'\right)\left(\bar{\pi}-\pi^0\right) \\ &\geq C_1\left\|\bar{\pi}-\pi^0\right\|^2/2 \text{ w.p.a.1.} \end{aligned} \tag{A.3}$$

Combining (A.2) and (A.3) yields

$$\left\|\bar{\pi}-\pi^0\right\|=O_p\left(\eta_{1Kn}\right). \tag{A.4}$$

Next,

$$\begin{aligned}
&\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|R(x)'\bar{\pi}-LF\left(\sum_{l=1}^{d}g_l(x_{l,t_1})-\sum_{l=1}^{d}g_l(x_{l,t_2})\right)\right| \\
\leq &\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|R(x)'\left(\bar{\pi}-\pi^0\right)\right|+\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|R(x)'\pi^0-LF\left(\sum_{l=1}^{d}g_l(x_{l,t_1})-\sum_{l=1}^{d}g_l(x_{l,t_2})\right)\right| \\
\leq &\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\|R(x)\|\left\|\bar{\pi}-\pi^0\right\|+O\left(K^{-\gamma}\right) \\
= &\ \zeta_{0K}O_p\left(\eta_{1Kn}\right)+O\left(K^{-\gamma}\right)=O_p\left(\zeta_{0K}\eta_{1Kn}\right). \qquad \text{(A.5)}
\end{aligned}$$

By (A.1) and the uniform boundedness of the first derivative of $L(\cdot)$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left[\bar{E}\left(D_i|X_i\right)-E\left(D_i|X_i\right)\right]^2 &\leq \frac{2}{n}\sum_{i=1}^{n}\left[\bar{E}\left(D_i|X_i\right)-L\left(R(X_i)'\pi^0\right)\right]^2+\frac{2}{n}\sum_{i=1}^{n}\left[r_L\left(X_i\right)\right]^2 \\
&= \frac{2}{n}\sum_{i=1}^{n}\left[L\left(R(X_i)'\bar{\pi}\right)-L\left(R(x)'\pi^0\right)\right]^2+\frac{2}{n}\sum_{i=1}^{n}\left[r_L\left(X_i\right)\right]^2 \\
&\lesssim \frac{2}{n}\sum_{i=1}^{n}\left[R(X_i)'\left(\bar{\pi}-\pi^0\right)\right]^2+\frac{2}{n}\sum_{i=1}^{n}\left[r_L\left(X_i\right)\right]^2 \\
&= O_p\left(\eta_{1Kn}^2\right)+O_p\left(K^{-2\gamma}\right)=O_p\left(\eta_{1Kn}^2\right), \qquad \text{(A.6)}
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|\bar{E}\left(D_i|X_i=x\right)-E\left(D_i|X_i=x\right)\right| \\
\leq &\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|\bar{E}\left(D_i|X_i=x\right)-L\left(R(x)'\pi^0\right)\right|+\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|r_L\left(x\right)\right| \\
\lesssim &\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|R(x)'\left(\bar{\pi}-\pi^0\right)\right|+\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}}\left|r_L\left(x\right)\right| \\
= &\ \zeta_{0K}O_p\left(\eta_{1Kn}\right)+O_p\left(K^{-\gamma}\right)=O_p\left(K^{1/2}\eta_{1Kn}\right). \qquad \text{(A.7)}
\end{aligned}$$

A.1.2 Convergence Rate of $\bar{g}_l(\cdot)$

Noting that $\triangle P_{i,j}^{1,K}=\triangle g_{i,j}-\theta_0{}'\triangle P_{i,j}^{K-1,K}+\left\{\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right\}$ 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 (2.12) we have

$$\begin{aligned}
\bar{\theta}-\theta_0=&-\left\{\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}\bar{H}_{1h_1,ji}\right\}^{-1}\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}\bar{H}_{1h_1,ji} \\
&-\left\{\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}\bar{H}_{1h_1,ji}\right\}^{-1}\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left\{\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right\}\triangle P_{i,j}^{K-1,K}\bar{H}_{1h_1,ji}
\end{aligned}$$

$$\equiv -L_{0,n}^{-1}L_{1,n} - L_{0,n}^{-1}L_{2,n}, \tag{A.8}$$

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

$$\begin{aligned} L_{1,n} &= \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle g_j \triangle P_{i,j}^{K-1,K} \bar{H}_{1h_1,ji} - \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle g_i \triangle P_{i,j}^{K-1,K} \bar{H}_{1h_1,ji} \\ &= \frac{2}{n^2}\sum_{1\leq i\neq j\leq n} \triangle g_j \triangle P_{i,j}^{K-1,K} \bar{H}_{1h_1,ji}. \end{aligned} \tag{A.9}$$

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

$$\begin{aligned} &\bar{H}_{1h_1,ji} - H_{1h_1,ji} \\ &=H_{h_1}\left[\bar{E}\left(D_j|X_j\right) - \bar{E}\left(D_j|X_i\right)\right] - H_{1h_1}\left[E\left(D_j|X_j\right) - E\left(D_j|X_i\right)\right] \\ &= h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_{ji} - m_{ji}\right) + \frac{1}{2}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_{ji} - m_{ji}\right)^2 + \frac{1}{6}h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)\left(\bar{m}_{ji} - m_{ji}\right)^3 \\ &= h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j - m_j\right) - h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i - m_i\right) \\ &- h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i - m_i\right)\left(\bar{m}_j - m_i\right) + \frac{1}{2}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j - m_j\right)^2 \\ &+ \frac{1}{2}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i - m_i\right)^2 + h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)\left(\bar{m}_{ji} - m_{ji}\right)^3, \end{aligned} \tag{A.10}$$

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

$$\begin{aligned} L_{0,n} &= \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1,ji} \\ &+ \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j - m_j\right) \\ &- \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i - m_i\right) \\ &- \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i - m_i\right)\left(\bar{m}_j - m_i\right) \\ &+ \frac{1}{2n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j - m_j\right)^2 \\ &+ \frac{1}{2n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i - m_i\right)^2 \\ &+ \frac{1}{6n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)\left(\bar{m}_{ji} - m_{ji}\right)^3 \\ &\equiv L_{0,n1} + L_{0,n2} + ... + L_{0,n7}. \end{aligned} \tag{A.11}$$

By B.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 $\lambda_{\min}(L_{0,n}) \geq c_{L_0}/2$ w.p.a.1.

Next, we derive the asymptotic properties of $L_{1,n}$. By (A.10), the symmetry of the kernel function $H(\cdot)$, and the fact that $\triangle P_{i,j}^{K-1,K} = -\triangle P_{j,i}^{K-1,K}$,

$$
\begin{aligned}
L_{1,n} &= \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle g_{i,j}\triangle P_{i,j}^{K-1,K} H_{1h_1,ji} + \frac{2}{n^2}\sum_{1\leq i\neq j\leq n} \triangle g_j \triangle P_{i,j}^{K-1,K} h_1^{-1}\dot{H}_{1h_1}(m_{ji})(\bar{m}_j - m_j) \\
&- \frac{2}{n^2}\sum_{1\leq i\neq j\leq n} \triangle g_j \triangle P_{i,j}^{K-1,K} h_1^{-1}\dot{H}_{1h_1}(m_{ji})(\bar{m}_i - m_i) \\
&- \frac{2}{n^2}\sum_{1\leq i\neq j\leq n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_i - m_i)(\bar{m}_j - m_i) \\
&+ \frac{1}{2n^2}\sum_{1\leq i\neq j\leq n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_j - m_j)^2 \\
&+ \frac{1}{2n^2}\sum_{1\leq i\neq j\leq n} \triangle g_j \triangle P_{i,j}^{K-1,K} h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_i - m_i)^2 \\
&+ \frac{1}{6n^2}\sum_{1\leq i\neq j\leq n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)(\bar{m}_{ji} - m_{ji})^3 \\
&\equiv \sum_{\ell=1}^{7} L_{1,n\ell}.
\end{aligned}
\tag{A.12}
$$

By Lemma B.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 (A.10),

$$
\begin{aligned}
L_{2,n} &= \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} H_{1h_1,ji} \\
&+ \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} h_1^{-1}\dot{H}_{1h_1}(m_{ji})[\bar{m}_j - m_j] \\
&- \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} h_1^{-1}\dot{H}_{1h_1}(m_{ji})[\bar{m}_i - m_i] \\
&- \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} h_1^{-2}\ddot{H}_{1h_1}(m_{ji})[\bar{m}_i - m_i][\bar{m}_j - m_i] \\
&+ \frac{1}{2n^2}\sum_{1\leq i\neq j\leq n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} h_1^{-2}\ddot{H}_{1h_1}(m_{ji})[\bar{m}_j - m_j]^2 \\
&+ \frac{1}{2n^2}\sum_{1\leq i\neq j\leq n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} h_1^{-2}\ddot{H}_{1h_1}(m_{ji})[\bar{m}_i - m_i]^2 \\
&+ \frac{1}{6n^2}\sum_{1\leq i\neq j\leq n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)(\bar{m}_{ji} - m_{ji})^3 \\
&\equiv \sum_{\ell=1}^{7} L_{2,n\ell}.
\end{aligned}
\tag{A.13}
$$

By Lemma B.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

$$\begin{aligned}
\bar{\theta} - \theta_0 = & -\left\{\frac{1}{n^2}\sum_{1\le i\ne j\le n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1,ji}\right\}^{-1} \\
& \times \left\{\frac{1}{n^2}\sum_{1\le i\ne j\le n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} H_{1h_1,ji} + \frac{1}{n^2}\sum_{1\le i\ne j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} H_{1h_1,ji}\right\} \\
& + R_{1n}, \qquad \text{(A.14)}
\end{aligned}$$

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

Given the result in (A.14) and using the results in B.1-B.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 (A.6)-(A.7), we have

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

and

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

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

A.2 Proof of Theorem 3.2

Let

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

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 $\left| l_{i,j}'''(\tau) \right| \leq \left| l_{i,j}''(\tau) \right| \left| \varsigma_{1i,j}' U_{\triangle g_j} \right|$. Define

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

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 $\tau = 0$ yields

$$
\partial_\tau Q_n \left(\triangle \bar{g}_j, (LF(\triangle g_j), \partial LF(\triangle g_j), \cdots, \partial^{a_2} LF(\triangle g_j))' + \tau U_{\triangle g_j} \right) \Big|_{\tau=0}
$$

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

Let

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

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 Bach (2010) 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 Jiang, Phillips, Tao and Zhang (2021), there exists a positive constant $\underline{c}^*$ such that

$$
\begin{aligned}
G_n\left(\widehat{U}_{\triangle g_j}\right)\geq&\frac{\underline{c}^*}{3}\min\left(\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{i,j}^2,\bar{l}\left[\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{i,j}^2\right]^{1/2}\right)\\
\geq&\frac{\underline{c}^*}{3}\min\left(\left\|\widehat{U}_{\triangle g_j}\right\|^2,\bar{l}\left\|\widehat{U}_{\triangle g_j}\right\|\right),
\end{aligned}
\tag{A.15}
$$

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

$$\bar{l}\left(h_2^{a_2+1}+\sqrt{\ln\left(n\right)/\left(nh_2\right)}\right)^{-1} \xrightarrow{p} \infty. \tag{A.16}$$

In addition, by construction and the submultiplicative and triangle inequalities

$$\begin{aligned}
G_n\left(\widehat{U}_{\triangle g_j}\right) &\leq \left|\partial_t Q_n\left(\triangle\bar{g}_j,(LF(\triangle g_j),\partial LF(\triangle g_j),\cdots,\partial^{a_2}LF(\triangle g_j))'+\tau U_{\triangle g_j}\right)\big|_{\tau=0}\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{1i,j}'\widehat{U}_{\triangle g_j}\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j)\right)-D_i\right\}\right| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{1i,j}\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j)\right)-D_i\right\}\right\|\left\|\widehat{U}_{\triangle g_j}\right\| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{1i,j}'\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j)\right)-F(\triangle g_i)\right\}\right\|\left\|\widehat{U}_{\triangle g_j}\right\| \\
&\quad+\left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{1i,j}'\left[F(\triangle g_i)-D_i\right]\right\|\left\|\widehat{U}_{\triangle g_j}\right\|.
\end{aligned} \tag{A.17}$$

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

$$\begin{aligned}
&L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j)\right)-F(\triangle g_i) \\
&=\left[F(\triangle\bar{g}_i)-F(\triangle g_i)\right]+\left[L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j)-LF(\triangle\bar{g}_i)+LF(\triangle\bar{g}_i)\right)-F(\triangle\bar{g}_i)\right] \\
&=\left[F(\triangle\bar{g}_i)-F(\triangle g_i)\right]+L'\left(LF(\triangle\bar{g}_i)\right)\chi_{i,j}+\frac{1}{2}L''\left(LF(\triangle\bar{\bar{g}}_{i,j})\right)\chi_{i,j}^2.
\end{aligned} \tag{A.18}$$

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 (A.18) into (A.17) yields

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

$$\equiv \{\|G_{1,nj}\| + \|G_{2,nj}\| + \|G_{3,nj}\| + \|G_{4,nj}\|\} \left\| \widehat{U}_{\triangle g_j} \right\|, \tag{A.19}$$

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

$$\begin{aligned} \left| G_n\left(\widehat{U}_{\triangle g_j}\right)\right| &\leq c_F \left|\triangle g_j - \triangle \bar{g}_j\right| \left\|\widehat{U}_{\triangle g_j}\right\| \\ &+ O_p\left(h_2^{a_2+1} + \sqrt{\log(n)/(nh_2)} + \eta_{2Kn}\right)\left\|\widehat{U}_{\triangle g_j}\right\|. \end{aligned} \tag{A.20}$$

Combining (A.15) and (A.20), we have uniformly over $j \in \{1, \cdots, n\}$

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

which, in conjunction with (A.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.

A.3 Proof of Theorem 3.3

A.3.1 Convergence Rate of $\left(\widehat{g}_l(x_l),\ \widehat{\dot{g}}_l(x_l)\right)'$

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

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

and

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

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

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

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

and similarly,

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

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

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

It is straightforward to show that

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

Define

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

We calculate the first order derivative at τ,

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

and

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

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

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

Let

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

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

$$\begin{aligned}G_n\left(U_{x_l}\right)&=\frac{1}{n}\sum_{i=1}^{N}\left[l_{l,i2}(1)-l_{l,i2}(0)-{l_{l,i2}}'(0)\right]+\frac{1}{n}\sum_{i=1}^{N}\left[l_{1,i}(1)-l_{1,i}(0)-{l_{1,i}}'(0)\right]\\ &\geq\frac{1}{n}\sum_{i=1}^{N}\frac{{l_i}''(0)}{\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i2}\right|^2}\left[\exp\left(-\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i2}\right|\right)+\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i2}\right|-1\right]\\ &\quad+\frac{1}{n}\sum_{i=1}^{N}\frac{{l_i}''(0)}{\left[\widehat{L\dot{F}}_i\cdot u_{x_l,i1}\right|^2}\left[\exp\left(-\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i1}\right|\right)+\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i1}\right|-1\right]\\ &=\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i2x_l}L'\left(\widehat{LF}_i-\widehat{L\dot{F}}_i\cdot r_{x_l,i2}\right)\left[\exp\left(-\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i2}\right|\right)+\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i2}\right|-1\right]\\ &\quad+\frac{1}{n}\sum_{i=1}^{N}T_{h_3}\left(X_{,l,i1}-x_l\right)L'\left(\widehat{LF}_i+\widehat{L\dot{F}}_i\cdot r_{x_l,i1}\right)\left[\exp\left(-\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i1}\right|\right)+\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i1}\right|-1\right]\\ &\geq C_1\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i2x_l}\left(\frac{1}{2}\left[\widehat{L\dot{F}}_i\cdot u_{x_l,i2}\right]^2-\frac{1}{6}\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i2}\right|^3\right)\\ &\quad+C_1\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i1x_l}\left(\frac{1}{2}\left[\widehat{L\dot{F}}_i\cdot u_{x_l,i1}\right]^2-\frac{1}{6}\left|\widehat{L\dot{F}}_i\cdot u_{x_l,i1}\right|^3\right),\end{aligned}$$

where the first inequality holds by Lemma 1 in Bach (2010) 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 Jiang et al. (2021), there exist some positive constant $\underline{C}_1$ and $\underline{C}$ such that

$$\begin{aligned}
G_n\left(\widehat{U}_{x_l}\right) \geq \frac{1}{3}\min\Bigg(&\underline{C}_1\frac{1}{n}\sum_{i=1}^{N} H_{3h_3,i1x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right]^2+\underline{C}_1\frac{1}{n}\sum_{i=1}^{N} H_{3h_3,i2x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right]^2,\\
&\underline{C}_1\bar{l}_1\left[\frac{1}{n}\sum_{i=1}^{N} H_{3h_3,i1x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right]^2\right]^{1/2}+\underline{C}_1\bar{l}_2\left[\frac{1}{n}\sum_{i=1}^{N} H_{3h_3,i2x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right]^2\right]^{1/2}\Bigg)\\
\geq&\frac{\underline{C}}{3}\min\left(\left\|\widehat{U}_{x_l}\right\|^2,\left(\bar{l}_1+\bar{l}_2\right)\left\|\widehat{U}_{x_l}\right\|\right),
\end{aligned}\tag{A.21}$$

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 (A.16), we have

$$\left(\bar{l}_1+\bar{l}_2\right)\left(h_3^2+\sqrt{1/(nh_3)}\right)^{-1}\xrightarrow{p}\infty.\tag{A.22}$$

In addition, by construction,

$$\begin{aligned}
G_n\left(\widehat{U}_{x_l}\right) &\leq \left|\partial_\tau W_{1,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)\big|_{\tau=0}+\partial_\tau W_{2,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)\big|_{\tau=0}\right|\\
&\leq\left|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i1x_l}u_{x_l,i1}\left\{L\left(\widehat{LF}_i+\widehat{\dot{LF}}_i\cdot r_{x_l,i1}\right)-D_i\right\}\right|\\
&+\left|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}u_{x_l,i2}\left\{L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i\cdot r_{x_l,i2}\right)-D_i\right\}\right|\\
&\leq\left\|U_{x_l}\right\|\left\{\left\|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i1x_l}\mu_{x_l,it}\left\{L\left(\widehat{LF}_i+\widehat{\dot{LF}}_i\cdot r_{x_l,i1}\right)-D_i\right\}\right\|\right.\\
&+\left.\left\|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\mu_{x_l,it}\left\{L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i\cdot r_{x_l,i2}\right)-D_i\right\}\right\|\right\}\\
&\equiv\left\|U_{x_l}\right\|\left\{\left\|D_{1n}\left(x_l\right)\right\|+\left\|D_{2n}\left(x_l\right)\right\|\right\},
\end{aligned}\tag{A.23}$$

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,

$$L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i\cdot r_{x_l,i2}\right)-D_i$$

$$= L\left(LF_i + \widehat{LF}_i - LF_i - \widehat{\dot{LF}}_i\left[\bar{g}_l\left(X_{l,i2}\right) - g_l\left(X_{l,i2}\right)\right] + \frac{1}{2}\widehat{\dot{LF}}_i\ddot{g}_l(\bar{x}_{li})\left(X_{l,i2} - x_l\right)^2\right) - D_i$$
$$= L\left(LF_i + \delta_{x_l,i2}\right) - D_i$$
$$= F_i - D_i + L'\left(LF_i\right)\delta_{x_l,i2} + \frac{1}{2}L''\left(\bar{LF}_i\right)\delta^2_{x_l,i2}, \tag{A.24}$$

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,

$$L\left(\widehat{LF}_i + \widehat{\dot{LF}}_i \cdot r_{x_l,i1}\right) - D_i$$
$$= L\left(LF_i + \widehat{LF}_i - LF_i + \widehat{\dot{LF}}_i\left[\bar{g}_l\left(X_{l,i1}\right) - g_l\left(X_{l,i1}\right)\right] - \frac{1}{2}\widehat{\dot{LF}}_i\ddot{g}_l(\bar{x}_{li})\left(X_{l,i1} - x_l\right)^2\right) - D_i$$
$$= L\left(LF_i + \delta_{x_l,i1}\right) - D_i$$
$$= F_i - D_i + L'\left(LF_i\right)\delta_{x_l,i1} + \frac{1}{2}L''\left(\bar{LF}_i\right)\delta^2_{x_l,i1}, \tag{A.25}$$

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 (A.24),

$$\begin{aligned}
D_{1n}\left(x_l\right) &\equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\mu_{x_l,it}\left\{L\left(\widehat{LF}_i - \widehat{\dot{LF}}_i \cdot r_{x_l,i2}\right) - D_i\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\dot{LF}_i H_{3h_3,i2x_l}\mu_{x_l,it}\left(F_i - D_i\right) + \frac{1}{N}\sum_{i=1}^{n}(\widehat{\dot{LF}}_i - \dot{LF}_i)H_{3h_3,i2x_l}\mu_{x_l,it}\left(F_i - D_i\right) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\mu_{x_l,it}L'\left(LF_i\right)\delta_{x_l,i2} + \frac{1}{2n}\sum_{i=1}^{n}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\mu_{x_l,it}L''\left(\bar{LF}_i\right)\delta^2_{x_l,i2} \\
&\equiv D_{1n,1}\left(x_l\right) + D_{1n,2}\left(x_l\right) + D_{1n,3}\left(x_l\right) + D_{1n,4}\left(x_l\right).
\end{aligned}$$

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 $D_{1n,4}\left(x_l\right) = O_p\left(h_3^4 + \eta_{3Kn}^2\right)$ uniformly in $x_l \in \mathcal{X}_l$ by using Theorem 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 (A.23)

$$G_n\left(\widehat{U}_{x_l}\right) \leq O_p\left(h_3^2 + \sqrt{1/(nh_3)} + \eta_{3Kn}\right)\left\|\widehat{U}_{x_l}\right\|. \tag{A.26}$$

Combining (A.21) and (A.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 (A.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: $\max_{x_l\in\mathcal{X}_l}\left\|\widehat{U}_{x_l}\right\| = O_p\left(h_3^2+\sqrt{\log(n)/(nh_3)}\right)$.

A.3.2 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 (A.25) and (A.24), we have

$$\left.\frac{\partial W_{1n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$$

$$= -\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i1x_l}\left(1, \frac{1}{h_3}(X_{l,i1}-x_l)\right)(F_i - D_i)$$

$$+ \frac{1}{2n}\sum_{i=1}^{N} L'(LF_i)\,\widehat{\dot{LF}}_i^2\,\ddot{g}_l(\bar{x}_{li})H_{3h_3,i1x_l}\left((X_{l,i1}-x_l)^2, \frac{1}{h_3}(X_{l,i1}-x_l)^3\right)$$

$$- \frac{1}{n}\sum_{i=1}^{N} L'(LF_i)\,\widehat{\dot{LF}}_i H_{3h_3,i1x_l}\left(1, \frac{1}{h_3}(X_{l,i1}-x_l)\right)\left\{\widehat{LF}_i - LF_i + \widehat{\dot{LF}}_i\left[\bar{g}_l(X_{l,i1}) - g_l(X_{l,i1})\right]\right\}$$

$$- \frac{1}{2n}\sum_{i=1}^{N} L''\left(\bar{LF}_i\right)\widehat{\dot{LF}}_i H_{3h_3,i1x_l}\left(1, \frac{1}{h_3}(X_{l,i1}-x_l)\right)\delta^2_{x_l,i1},$$

and

$$\left.\frac{\partial W_{2n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(a,b)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$$

$$= \frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}(X_{l,i2}-x_l)\right)(F_i - D_i)$$

$$+ \frac{1}{2n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}(X_{l,i2}-x_l)\right)L'(LF_i)\,\widehat{\dot{LF}}_i\ddot{g}_l(\bar{x}_{li})(X_{l,i2}-x_l)^2$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}(X_{l,i2}-x_l)\right)L'(LF_i)\left\{\widehat{LF}_i - LF_i - \widehat{\dot{LF}}_i\left[\bar{g}_l(X_{l,i2}) - g_l(X_{l,i2})\right]\right\}$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}(X_{l,i2}-x_l)\right) + \frac{1}{2}L''\left(\bar{LF}_i\right)\delta^2_{x_l,i2}.$$

Given the above results, we can show that

$$\sqrt{nh_3}\left(\frac{\partial W_{n,x_l}(g_l(x_l), h_3\dot{g}_l(x_l))}{\partial (c_0,c_1)'} - E\left[L'(LF(\triangle g(X_i)))\,\dot{LF}^2(\triangle g(X))\Big|\, X_{l,it}=x_l\right]\ddot{g}_l(x_l)f_{X_l}(x_l)\begin{pmatrix} h_3^2\kappa_{21} \\ 0\end{pmatrix}\right)$$

$$\xrightarrow{d} N\left(0, 2E\left\{\dot{LF}^2(\triangle g(X_i))F(\triangle g(X_i))\left[1-F(\triangle g(X_i))\right]\Big|\, X_{l,it}=x_l\right\}f_{X_l}(x_l)\begin{pmatrix}\kappa_{02} & 0 \\ 0 & \kappa_{22}\end{pmatrix}\right).$$

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:

$$\left.\frac{\partial^2 W_{n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'\partial (c_0,c_1)}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$$

$$= \left.\frac{\partial^2 W_{1,nx_l}(c_0,c_1)}{\partial (c_0,c_1)'\partial (c_0,c_1)}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))} + \left.\frac{\partial^2 W_{2,nx_l}(c_0,c_1)}{\partial (c_0,c_1)'\partial (c_0,c_1)}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$$

$$= \frac{1}{n}\sum_{i=1}^{N}\left(\widehat{\dot{LF}}_i\right)^2 H_{3h_3,i1x_l}\mu_{x_l,i1}L'\left(\widehat{LF}_i - \widehat{\dot{LF}}_i\cdot\left[g_l(x_l) + h_3\dot{g}_l(x_l)\frac{1}{h_3}(X_{l,i1}-x_l) - \bar{g}_l(X_{l,i1})\right]\right)$$

$$+\frac{1}{n}\sum_{i=1}^{N}\left(\widehat{\dot{LF}_i}\right)^2 H_{3h_3,i2x_l}\mu_{x_l,i2}L'\left(\widehat{LF}_i+\widehat{\dot{LF}_i}\cdot\left[g_l(x_l)+h_3\dot{g}_l(x_l)\frac{1}{h_3}\left(X_{l,i2}-x_l\right)-\bar{g}_l\left(X_{l,i2}\right)\right]\right)$$

$$\xrightarrow{p} 2E\left[L'\left(LF(\triangle g(X))\right)\dot{LF}^2(\triangle g(X))\Big|\, X_l=x_l\right]f_{X_l}(x_l)\begin{pmatrix}1 & 0\\ 0 & \kappa_{21}\end{pmatrix}$$

$$=2E\left[L'\left(LF(\triangle g(X))\right)\dot{LF}^2(\triangle g(X))\Big|\, X_l=x_l\right]f_{X_l}(x_l)\begin{pmatrix}1 & 0\\ 0 & \kappa_{21}\end{pmatrix},$$

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

$$\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)'-\left(g_l(x_l),\ h_3\dot{g}_l(x_l)\right)'$$

$$=\left(\left.\frac{\partial^2 W_{n,x_l}\left(c_0,c_1\right)}{\partial\left(c_0,c_1\right)'\partial\left(c_0,c_1\right)}\right|_{(c_0,c_1)=\left(g_l^*(x_l),\ h_3\dot{g}_l^*(x_l)\right)'}\right)^{-1}\left.\frac{\partial W_{n,x_l}\left(c_0,c_1\right)}{\partial\left(c_0,c_1\right)'}\right|_{(c_0,c_1)=\left(g_l(x_l),\ h_3\dot{g}_l(x_l)\right)'}$$

$$=\left(\frac{1}{n}\sum_{i=1}^{N}\left(\dot{LF}_i\right)^2 H_{3h_3,i1x_l}\mu_{x_l,it}L'\left(LF_i\right)\right)^{-1}$$

$$\times\left\{-\frac{1}{n}\sum_{i=1}^{N}\dot{LF}_i H_{3h_3,i1x_l}\left(1,\frac{1}{h_3}\left(X_{l,i1}-x_l\right)\right)\left(F_i-D_i\right)\right.$$

$$+\frac{1}{2n}\sum_{i=1}^{N}L'\left(LF_i\right)\dot{LF}_i^2\ddot{g}_l(\bar{x}_{li})H_{3h_3,i1x_l}\left(\left(X_{l,i1}-x_l\right)^2,\frac{1}{h_3}\left(X_{l,i1}-x_l\right)^3\right)$$

$$+\frac{1}{n}\sum_{i=1}^{N}\dot{LF}_i H_{3h_3,i2x_l}\left(1,\frac{1}{h_3}\left(X_{l,i2}-x_l\right)\right)\left(F_i-D_i\right)$$

$$\left.+\frac{1}{2n}\sum_{i=1}^{N}\dot{LF}_i H_{3h_3,i2x_l}\left(1,\frac{1}{h_3}\left(X_{l,i2}-x_l\right)\right)L'\left(LF_i\right)\dot{LF}_i\ddot{g}_l(\bar{x}_{li})\left(X_{l,i2}-x_l\right)^2\right\}+R_{3n},$$

where $\|R_{3n}\|=O_p\left(h_3^2+\sqrt{h_3/n}+\eta_{3Kn}\right)$. Then

$$\sqrt{nh_3}\left(\begin{pmatrix}1 & 0\\ 0 & h_3\end{pmatrix}\left(\begin{pmatrix}\widehat{g}_l(x_l)\\ \widehat{\dot{g}}_l(x_l)\end{pmatrix}-\begin{pmatrix}g_l(x_l)\\ \dot{g}_l(x_l)\end{pmatrix}\right)-\frac{1}{2}\ddot{g}_l(x_l)\begin{pmatrix}h_3^2\kappa_{21}\\ 0\end{pmatrix}\right)$$

$$\xrightarrow{d} N\left(0,\frac{1}{2}\left(E\left\{\dot{LF}^2(\triangle g(X))F(\triangle g(X))\left[1-F(\triangle g(X))\right]\Big|\, X_l=x_l\right\}f_{X_l}(x_l)\right)^{-1}\begin{pmatrix}\kappa_{02} & 0\\ 0 & \frac{\kappa_{22}}{\kappa_{21}^2}\end{pmatrix}\right),$$

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 $\left(\widehat{g}_l(x_l)\ \widehat{\dot{g}}_l(x_l)\right)$,

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

This completes the proof of the theorem.

B 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.1 *Let $L_{0,n}$ and $L_{0,n1}$ be as defined in the proof of Theorem 3.1. Suppose that the conditions in Theorem 3.1 are satisfied. Then*

(i) $\|L_{0,n1}-E\left(L_{0,n1}\right)\| = O_p\left(\sqrt{K/n}\right)$, $\|E\left(L_{0,n1}\right)-E\left[\eta(m_i)f_m\left(m_i\right)\right]\| = O\left(Kh^{a_1}\right)=o\left(1\right)$, *and* $\lambda_{\min}\left(L_{0,n1}\right)\geq C_{1L}/2$ *w.p.a.1.*;

(ii) $\left\|L_{0,n\ell}\right\|_{op}=O_p\left(K^{1/2}h_1^{-1}\eta_{1Kn}\right)=o_p(1)$ *for* $\ell=2,...,7$;

(iii) $\left\|L_{0,n}-E\left[\eta(m_i)f_m\left(m_i\right)\right]\right\|_{op}=o_p(1)$ *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 $j\neq i$

$$\begin{aligned}E\left[\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}H_{1h_1}(m_j-m_i)\right] &= E\left\{\eta(m_i)H_{1h_1}(m_j-m_i)\right\}\\ &= E\left[\eta(m_i)\int\frac{1}{h_1}H_1(\frac{m-m_i}{h_1})f_m\left(m\right)dm\right]\\ &= E\left[\eta(m_i)\int H_1(u)f_m\left(m_i+hu\right)du\right]\\ &= E\left[\eta(m_i)f_m\left(m_i\right)\right]+O\left(h_1^{a_1}\right)E\left[\eta(m_i)f_m^{(a_1)}\left(m_i\right)\right],\end{aligned}$$

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,

$$\lambda_{\min}\left(L_{0,n1}\right)\ \geq\ \lambda_{\min}\left(E\left(L_{0,n1}\right)\right)-\left\|L_{0,n1}-E\left(L_{0,n1}\right)\right\|$$

$$\geq \lambda_{\min}\left(E\left(L_{0,n1}\right)\right)-O_p\left(\sqrt{K/n}\right)$$
$$\geq \frac{1}{2}\lambda_{\min}\left(E\left[\eta(m_i)f_m\left(m_i\right)\right]\right)-O\left(Kh^{a_1}\right)-O_p\left(\sqrt{K/n}\right)$$
$$\geq C_{1L}/2 \text{ w.p.a.1.}$$

(ii) As in part (i), we can readily show that $\|L_{0,n\ell}\|_{\text{op}} = O_p\left(K^{1/2}h_1^{-1}\eta_{1Kn}\right) = o_p(1)$ for $\ell = 2, ..., 7$. For example, for $L_{0,n2}$, we have

$$\begin{aligned}\|L_{0,n2}\|_{op} &= \left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}{h_1}^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j-m_j\right)\right\|_{op}\\ &\leq \max_{1\leq j\leq n}\left|\bar{m}_j-m_j\right|\frac{1}{n^2}\sum_{j=1}^{n}\left\|\sum_{i=1,i\neq j}^{n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}{h_1}^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\right\|_{op}\\ &= O_p\left(K^{1/2}\eta_{1Kn}\right)O_p\left(h_1^{-1}\right)=O_p\left(K^{1/2}h_1^{-1}\eta_{1Kn}\right)=o_p\left(1\right).\end{aligned}$$

(iii) The result follows from (i)-(ii) and the Weyl's inequality. ■

Lemma B.2 *Let $L_{1,n}$, $L_{1,n1}$ and $L_{1,n2}$ be as defined in the proof of Theorem 3.1. Suppose that the conditions in Theorem 3.1 are satisfied. Then*
(i) $\|L_{1,n1}\| = O_p\left(\sqrt{K/n}+\sqrt{K}h_1^{a_1}\right)$;
(ii) $\|L_{1,n\ell}\| = O_p\left(\eta_{1Kn}\right)$ *for* $\ell = 2, 3$;
(iii) $\|L_{1,n\ell}\| = O_p\left(\eta_{1Kn}\right)$ *for* $\ell = 4, 5, 6, 7$;
(iv) $\|L_{1,n}\| = O_p\left(\eta_{1Kn}+\sqrt{K}h_1^{a_1}\right)$.

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 $q_{1n}(X_i, X_j) = \triangle g_{ij}\triangle P_{i,j}^{K-1,K}H_{1h_1}\left(m_{ji}\right)$. 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

$$\mathbb{U}_{1n}^{(1)} = \frac{1}{n}\sum_{i=1}^{n}\left[r_{1n}(X_j)-\theta_{1n}\right],$$

$$\mathbb{U}_{1n}^{(2)} = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} [q_{1n}(X_i, X_j) - r_{1n}(X_i) - r_{1n}(X_j) + \theta_{1n}].$$

Note that for $j \neq i$,

$$E\|q_{1n}(X_i, X_j)\|^2 = E\left[\mathrm{tr}\left\{(\triangle g_{i,j})^2 \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1}^2(m_{ji})\right\}\right] = O(Kh^{-1}) = o(n).$$

Then by Lemma 3.1 in Powell, Stock and Stoker (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 $\rho_P(m_i) = E\left[\triangle P_i^{K-1,K}|m_i\right]$, $\rho_g(m_i) = E[\triangle g_i|m_i]$ and $\rho(m_i) = E\left[\triangle g_i \triangle P_i^{K-1,K}|m_i\right]$ Note that $r_{1n}(X_j) = E\left[(\triangle g_i - \triangle g_j)\triangle P_{i,j}^{K-1,K} H_{1h_1}(m_{ji})|X_j\right] = E\left[\triangle g_i \triangle P_{i,j}^{K-1,K} H_{1h_1}(m_{ji})|X_j\right] - \triangle g_j E\left[\triangle P_{i,j}^{K-1,K} H_{1h_1}(m_{ji})|X_j\right] \equiv r_{1n,1}(X_j) - r_{1n,2}(X_j)$. By straightforward moment calculations and the independence of $\{X_i\}$, we have

$$\begin{aligned}
r_{1n,1}(X_j) &= E\left\{\triangle g_i \left[\triangle P_j^{K-1,K} - \triangle P_i^{K-1,K}\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle P_j^{K-1,K} E[E(\triangle g_i|m_i, X_j) H_{1h_1}(m_{ji})|X_j] \\
&- E\left\{E\left[\triangle g_i \triangle P_i^{K-1,K}(X_i)|m_i, X_j\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle P_j^{K-1,K} E[\rho_g(m_i) H_{1h_1}(m_{ji})|X_j] - E\{\rho(m_i) H_{1h_1}(m_{ji})|X_j\} \\
&= \triangle P_j^{K-1,K} \int \rho_g(m) \frac{1}{h_1} H_1\left(\frac{m - m_j}{h_1}\right) f_m(m)\,dm - \int \rho(m) \frac{1}{h_1} H_1\left(\frac{m - m_j}{h_1}\right) f_m(m)\,dm \\
&= \triangle P_j^{K-1,K}\left[\rho_g(m_j) f_m(m_j) + \frac{h_1^{a_1}}{a_1!}\partial^{(a_1)}[\rho_g(m) f_m(m)]\Big|_{m=m_j} + o(h_1^{a_1})\right] \\
&- \left[\rho(m_j) f_m(m_j) + \frac{h_1^{a_1}}{a_1!}\partial^{(a_1)}[\rho(m) f_m(m)]\Big|_{m=m_j} + o(h_1^{a_1})\right].
\end{aligned}$$

and

$$\begin{aligned}
r_{1n,2}(X_j) &= \triangle g_j E\left\{\left[\triangle P_j^{K-1,K} - \triangle P_i^{K-1,K}\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle g_j \triangle P_j^{K-1,K} E[H_{1h_1}(m_{ji})|X_j] - \triangle g_j E\left\{E\left[\triangle P_i^{K-1,K}|m_i, X_j\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle g_j \triangle P_j^{K-1,K} E[H_{1h_1}(m_{ji})|X_j] - \triangle g_j E\{\rho_P(m_i) H_{1h_1}(m_{ji})|X_j\} \\
&= \triangle g_j \triangle P_j^{K-1,K} \int \frac{1}{h_1} H_1\left(\frac{m - m_j}{h_1}\right) f_m(m)\,dm - \triangle g_j \int \rho_P(m) \frac{1}{h_1} H_1\left(\frac{m - m_j}{h_1}\right) f_m(m)\,dm \\
&= \triangle g_j \triangle P_j^{K-1,K}\left[f_m(m_j) + \frac{h_1^{a_1}}{a_1!} f_m^{(a_1)}(m_j) + o(h_1^{a_1})\right] \\
&- \triangle g_j \left[\rho(m_j) f_m(m_j) + \frac{h_1^{a_1}}{a_1!}\partial^{(a_1)}[\rho_P(m) f_m(m)]\Big|_{m=m_j} + o(h_1^{a_1})\right].
\end{aligned}$$

Then it is easy to show that

$$\|\theta_{1n}\| = \|E[r_{1n,1}(X_j)] - E[r_{1n,2}(X_j)]\| = O\left(\sqrt{K} h_1^{a_1}\right)$$

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 $E\left\|r_{1n}(X_j)\right\|^2 = O(K)$ 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

$$\begin{aligned}
q_{2n}\left((X_i, X_j), \pi\right) & = h_1^{-1} \triangle g_j \triangle P_{i,j}^{K-1,K} \dot{H}_{1h_1}(m_{ji})\left[L\left(R(X_j)'\pi\right) - m_j\right] \\
& \quad + h_1^{-1} \triangle g_i \triangle P_{j,i}^{K-1,K} \dot{H}_{1h_1}(m_{ij})\left[L\left(R(X_i)'\pi\right) - m_i\right] \\
& = \left\{\triangle g_j\left[L\left(R(X_j)'\pi\right) - m_j\right] + \triangle g_i\left[L\left(R(X_i)'\pi\right) - m_i\right]\right\} h_1^{-1} \triangle P_{i,j}^{K-1,K} \dot{H}_{1h_1}(m_{ij}), \\
U_{2n}(\pi) & = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} q_n\left((X_i, X_j), \pi\right),
\end{aligned}$$

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, $q_{2n}\left((X_i, X_j), \pi\right)$ is symmetric in (X_i, X_j).

It suffices to determine the probability order of $U_{2n}(\bar{\pi})$ by studying the U-process $\{U_n(\pi)\}$. Let $r_{2n}(X_j,\pi) = E[q_{2n}((X_i,X_j),\pi)|X_j]$ and $\theta_{2n}(\pi) = E[r_{2n}(X_j,\pi)]$. 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

$$\begin{aligned}
\mathbb{U}_{2n}^{(1)}(\pi) &= \frac{1}{n}\sum_{i=1}^{n}[r_{2n}(X_j,\pi) - \theta_{2n}(\pi)] \\
\mathbb{U}_{2n}^{(2)}(\pi) &= \frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}[q_{2n}((X_i,X_j),\pi) - r_{2n}(X_j,\pi) - r_{2n}(X_j,\pi) + \theta_{2n}(\pi)].
\end{aligned}$$

Let $\delta_j \equiv \delta_{j,\pi} \equiv L(R(X_j)'\pi) - m_j$, 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

$$\begin{aligned}
r_{2n,2}(X_j,\pi) &= h_1^{-1}E\left[\triangle g_i\delta_i\triangle P_i^{K-1,K}\dot{H}_{1h_1}(m_{ij})|X_j\right] - h_1^{-1}\triangle P_j^{K-1,K}E\left[\triangle g_i\delta_i\dot{H}_{1h_1}(m_{ij})|X_j\right] \\
&= h_1^{-1}E\left[\rho_\delta(m_i)\dot{H}_{1h_1}(m_{ij})|X_j\right] - h_1^{-1}\triangle P_j^{K-1,K}E\left[\rho_{\delta g}(m_i)\dot{H}_{1h_1}(m_{ij})|X_j\right] \\
&= h_1^{-1}\int\dot{H}_1(u)\rho_\delta(m_j+hu)f_m(m_j+hu)\,du \\
&\quad - h_1^{-1}\triangle P_j^{K-1,K}\int\dot{H}_1(u)\rho_{\delta g}(m_j+hu)f_m(m_j+hu)\,du \\
&= -\partial\left[\rho_\delta(m_j)f_m(m_j)\right] + \triangle P_j^{K-1,K}\partial\left[\rho_{\delta g}(m_j)f_m(m_j)\right] + r_{2n2,a}(X_j,\pi),
\end{aligned}$$

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,

$$\begin{aligned}
\delta_j &= L\left(R(X_j)'\pi\right) - E(D_j|X_j) \\
&= L\left(R(X_j)'\pi\right) - L(LF(\triangle g_j)) \\
&= \dot{L}(LF(\triangle g_j))\left[R(X_j)'\pi - LF(\triangle g_j)\right] + \frac{1}{2}\ddot{L}(LF(\triangle\bar{\bar{g}}_j))\left[R(X_j)'\pi - LF(\triangle g_j)\right]^2 \\
&= \dot{L}(LF(\triangle g_j))\,R(X_j)'(\pi-\pi^0) + \dot{L}(LF(\triangle g_j))\left[R(X_j)'\pi^0 - LF(\triangle g_j)\right] \\
&\quad + \frac{1}{2}\ddot{L}(LF(\triangle\bar{\bar{g}}_j))\left\{R(X_j)'\left(\pi-\pi^0\right) + \left[R(X_j)'\pi^0 - LF(\triangle g_j)\right]\right\}^2 \\
&\equiv \dot{L}(LF(\triangle g_j))\,R(X_j)'\left(\pi-\pi^0\right) + \dot{L}(LF(\triangle g_j))\left[R(X_j)'\pi^0 - LF(\triangle g_j)\right] + \delta_{j,2},
\end{aligned}$$

where $\triangle\bar{\bar{g}}_j$ is between $R(X_j)'\pi$ and $LF(\triangle g_j)$. Note that

$$\begin{aligned}
\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}}\max_{1\le j\le n}|\delta_{j,2}| &\lesssim \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}}\left\|R(X_j)'\left(\pi-\pi^0\right)\right\|^2 \\
&\quad + \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}}\left\|R(X_j)'\pi^0 - LF(\triangle g_j)\right\|^2 \\
&\le K\left\|\pi-\pi^0\right\|^2 + O\left(K^{-2\gamma}\right) = O\left(K\eta_{1Kn}^2\right), \\
\max_{1\le j\le n}\left|R(X_j)'\pi^0 - LF(\triangle g_j)\right| &= O\left(K^{-\gamma}\right), \text{ and} \\
\sup_{\|\pi-\pi^0\|\le C\eta_{Kn}}\max_{1\le j\le n}\left|LF(\triangle\bar{\bar{g}}_j) - LF(\triangle g_j)\right| &\le O\left(\sqrt{K}\eta_{1Kn}\right).
\end{aligned}$$

In addition,

$$\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} E\left(\delta_j^2\right) \lesssim \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} E\left|R(X_j)'\left(\pi-\pi^0\right)\right|^2 + E\left[R(X_j)'\pi^0 - LF(\triangle g_j)\right]^2$$

$$
\begin{aligned}
&+ \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \max_{1\leq j\leq n} |\delta_{j,2}|^2 \\
= \ & O\left(\eta_{1Kn}^2\right) + O\left(K^{-2\gamma_F}\right) + O\left(K^2\eta_{1Kn}^4\right) = O\left(\eta_{1Kn}^2\right).
\end{aligned}
$$

By these results and the uniform boundedness of $\dot{L}$ and $\ddot{L}$, we can readily show that for $l = 1, 2$,

$$
\begin{aligned}
E\left[f_m(m_j)\psi_{1j}\right] &= E\left\{\left[\rho_{\delta g}(m_j)\rho_P^{(1)}(m_j) - \rho_P(m_j)\rho_{\delta g}^{(1)}(m_j) + \rho_\delta^{(1)}(m_j)\right] f_m(m_j)\right\} = O(\eta_{1Kn}), \\
\left|E\left[f_m^{(l)}\left(m_j^*\right)\psi_{1j}\right]\right| &\lesssim E\left\{\left|\rho_{\delta g}(m_j)\rho_P^{(1)}(m_j) - \rho_P(m_j)\rho_{\delta g}^{(1)}(m_j) + \rho_\delta^{(1)}(m_j)\right|\right\} = O\left(K^{1/2}\eta_{1Kn} + K\eta_{1Kn}^2\right),
\end{aligned}
$$

uniformly in π with $\left\|\pi - \pi^0\right\| \leq C\eta_{1Kn}$. Then

$$
\begin{aligned}
& \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} |\theta_{2n,1}(\pi)| \\
= \ & \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left|\int \dot{H}_1(u)\, uE\left[f_m(m_j + hu)\psi_{1j}\right] du\right| \\
= \ & \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left|-E\left[f_m(m_j)\psi_{1j}\right] + h_1\int \dot{H}_1(u)\, u^2 E\left[f_m^{(1)}(m_j)\psi_{1j}\right] du\right. \\
& \left. + \frac{h_1^2}{2}\int \dot{H}_1(u)\, u^3 du E\left[f_m^{(2)}\left(m_j^*\right)\psi_{1j}\right]\right| \\
\lesssim \ & \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left|E\left[f_m(m_j)\psi_{1j}\right]\right| + \frac{h_1^2}{2}\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left|E\left[f_m^{(2)}\left(m_j^*\right)\psi_{1j}\right]\right| \\
= \ & O(\eta_{1Kn}) + h_1^2 O\left(K^{1/2}\eta_{1Kn} + K\eta_{1Kn}^2\right) = O(\eta_{1Kn}),
\end{aligned}
$$

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 $\int \dot{H}_1(u)\, u^2 du = 0$ by the symmetry of $H_1(\cdot)$. Analogously, we have uniformly in π with $\left\|\pi - \pi^0\right\| \leq C\eta_{1Kn}$,

$$
\begin{aligned}
\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} |\theta_{2n,2}(\pi)| &= \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left|\frac{h_1}{2}\int \dot{H}_1(u)\, u^2 E\left[f_m(m_j + hu)\psi_{2j}\right] du\right| \\
&\leq \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \frac{h_1^2}{2}\left|\int \dot{H}_1(u)\, u^3 du E\left[f_m^{(1)}\left(m_j^*\right)\psi_{2j}\right]\right| \\
&= h_1^2 O\left(K^{1/2}\eta_{1Kn} + K\eta_{1Kn}^2\right) = O(\eta_{1Kn}).
\end{aligned}
$$

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

$$
E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|q_{2n}\left((X_1, X_2), \pi^0\right)\right\|^2\right]
$$

$$
\begin{aligned}
&\lesssim E\left[\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\| \triangle g_1 \left[L\left(R(X_1)'\pi^0\right) - m_1\right] h_1^{-1} \triangle P_{1,2}^{K-1,K} \dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
&\lesssim E\left[\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\| \left\{\triangle g_1 \left[L\left(R(X_1)'\pi^0\right) + L\left(R(X_1)'\pi^0\right)\right]\right\} h_1^{-1} \triangle P_{1,2}^{K-1,K} \dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
&\quad +E\left[\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\| \left\{\triangle g_1 \left[L\left(R(X_1)'\pi^0\right) - m_1\right]\right\} h_1^{-1} \triangle P_{1,2}^{K-1,K} \dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
&\lesssim E\left[\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\| \left\{\triangle g_1 R(X_1)'\left(\pi^0 - \pi^0\right)\right\} h_1^{-1} \triangle P_{1,2}^{K-1,K} \dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
&\quad +E\left[\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\| \left\{\triangle g_1 \left[L\left(R(X_1)'\pi^0\right) - m_1\right]\right\} h_1^{-1} \triangle P_{1,2}^{K-1,K} \dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
&\lesssim {h_1}^{-3} K \eta_{1Kn}^2 .
\end{aligned}
$$

Then by Corollary 5.3 in Chen and Kato (2020), we have

$$
\sup_{\|\pi-\pi^0\|\le C\eta_{Kn}} \left\| \mathbb{U}_{2n}^{(2)}(\pi)\right\| \lesssim n^{-1}\left({h_1}^{-3} K \eta_{1Kn}^2\right)^{1/2} \ln(n) = o_p\left(\eta_{1Kn}\right).
$$

By the empirical process theory, we have

$$
\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\| \mathbb{U}_{2n}^{(1)}(\pi)\right\| = O_p\left(n^{-1/2} K^{1/2} \ln(n)\right) = O_p\left(\eta_{1Kn}\right).
$$

Consequently, we have

$$
\left\|U_n\left(\bar{\pi}\right)\right\| \lesssim \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\| \theta_{2n}(\pi) + 2\mathbb{U}_{2n}^{(1)}(\pi) + \mathbb{U}_{2n}^{(2)}(\pi)\right\| = O_p\left(\eta_{1Kn}\right)
$$

and $\|L_{1,n2}\| = O_p\left(\eta_{1Kn}\right)$. Analogously, we can show that $\|L_{1,n3}\| = O_p\left(\eta_{1Kn}\right)$.

(iii) It suffices to obtain the rough probability bound for $\|L_{1,n\ell}\|$ with $\ell = 4, 5, 6, 7$. For example,

$$
\begin{aligned}
\|L_{1,n5}\| &\lesssim \left\| h_1^{-2} \frac{1}{n^2} \sum_{1\le i\ne j\le n} \triangle g_j \triangle P_{i,j}^{K-1,K} \ddot{H}_{1h_1,ji} \left| L\left(R(X_j)'\bar{\pi}\right) - E\left(D_j|X_j\right)\right|^2 \right\| \\
&\lesssim h_1^{-2} \frac{1}{n^2} \sum_{1\le i\ne j\le n} \left\| \triangle g_j \triangle P_{i,j}^{K-1,K} \ddot{H}_{1h_1,ji}\right\| \left| \dot{L}\left(LF\left(\triangle g_j\right)\right) R(X_j)'\left(\bar{\pi} - \pi^0\right)\right|^2 \\
&\quad + h_1^{-2} \frac{1}{n^2} \sum_{1\le i\ne j\le n} \left\| \triangle g_j \triangle P_{i,j}^{K-1,K} \ddot{H}_{1h_1,ji}\right\| \left| \left[R(X_j)'\pi^0 - LF\left(\triangle g_j\right)\right]\right|^2 \\
&\quad + h_1^{-2} \frac{1}{n^2} \sum_{1\le i\ne j\le n} \left\| \triangle g_j \triangle P_{i,j}^{K-1,K} \ddot{H}_{1h_1,ji}\right\| \left|\delta_{j,2}\right|^2 \\
&= h_1^{-2} O_p\left(K^{1/2}\eta_{1Kn}^2\right) + h_1^{-2} O_p\left(K^{-2\gamma}\right) + h_1^{-2} O_p\left(K^2 \eta_{1Kn}^4\right) = O_p\left(\eta_{1Kn}\right).
\end{aligned}
$$

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 $L_{1,n4} = \dfrac{1}{2n^2} \sum_{1 \le i \ne j \le n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} h_1^{-2} \ddot{H}_{1h_1}(m_{ji}) [\bar{m}_i - m_i][\bar{m}_j - m_j]$. It is easy to see that

$$\begin{aligned} L_{1,n4} &= \frac{1}{2n^2} \sum_{1 \le i \ne j \le n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} h_1^{-2} \ddot{H}_{1h_1}(m_{ji}) [\bar{m}_i - m_i][\bar{m}_j - m_j] \\ &= \frac{1}{2n^2} \sum_{1 \le i < j \le n} q_{3n}((X_i, X_j), \bar{\pi}) = \frac{1}{2} \varphi_n U_{3n}(\bar{\pi}) \end{aligned}$$

where

$$\begin{aligned} q_{3n}((X_i, X_j), \pi) &= \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} h_1^{-2} \ddot{H}_{h_1}(m_{ji}) \left[L\left(R(X_i)'\pi\right) - m_i\right] \left[L\left(R(X_j)'\pi\right) - m_j\right] \\ U_{3n}(\pi) &= \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} q_{3n}((X_i, X_j), \pi), \end{aligned}$$

and $\varphi_n = \binom{n}{2}/n^2 \to 1$ as $n \to \infty$. It suffices to determine the probability order of $U_{3n}(\bar{\pi})$ by studying the U-process $\{U_{3n}(\pi)\}$. Let $r_{3n}(X_j, \pi) = E[q_{3n}((X_i, X_j), \pi) | X_j]$ and $\theta_{3n}(\pi) = E[r_{3n}(X_j, \pi)]$. Then we have the following Hoeffding decomposition:

$$U_{3n}(\pi) = \theta_{3n}(\pi) + 2\mathbb{U}_{3n}^{(1)}(\pi) + \mathbb{U}_{3n}^{(2)}(\pi),$$

where

$$\begin{aligned} \mathbb{U}_{3n}^{(1)}(\pi) &= \frac{1}{n} \sum_{i=1}^{n} [r_{3n}(X_j, \pi) - \theta_{3n}(\pi)] \\ \mathbb{U}_{3n}^{(2)}(\pi) &= \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} [q_{3n}((X_i, X_j), \pi) - r_{3n}(X_j, \pi) - r_{3n}(X_j, \pi) + \theta_{3n}(\pi)]. \end{aligned}$$

Note that

$$\begin{aligned} r_{3n}(X_j, \pi) &= E\left\{\triangle g_{i,j} \triangle P_{i,j}^{K-1,K} h_1^{-2} \ddot{H}_{1h_1}(m_{ji}) \left[L\left(R(X_i)'\pi\right) - m_i\right] \left[L\left(R(X_j)'\pi\right) - m_j\right] | X_j\right\} \\ &= h_1^{-2} \delta_j E\left[\triangle g_{i,j} \delta_i \triangle P_{i,j}^{K-1,K} \ddot{H}_{1h_1}(m_{ji}) | X_j\right] \\ &= h_1^{-2} \delta_j E\left[\triangle g_i \delta_i \triangle P_{i,j}^{K-1,K} \ddot{H}_{1h_1}(m_{ji}) | X_j\right] - h_1^{-2} \delta_j \triangle g_j E\left[\delta_i \triangle P_{i,j}^{K-1,K} \ddot{H}_{1h_1}(m_{ji}) | X_j\right] \\ &\equiv r_{3n,1}(X_j, \pi) - r_{3n,2}(X_j, \pi). \end{aligned}$$

Note that

$$\begin{aligned} r_{3n,1}(X_j, \pi) &= h_1^{-2} \delta_j E\left[\triangle g_i \delta_i \triangle P_i^{K-1,K} \ddot{H}_{1h_1}(m_{ji}) | X_j\right] - h_1^{-2} \delta_j \triangle P_j^{K-1,K} E\left[\triangle g_i \delta_i \ddot{H}_{1h_1}(m_{ji}) | X_j\right] \\ &= h_1^{-2} \delta_j E\left[\rho(m_i) \ddot{H}_{1h_1}(m_{ij}) | X_j\right] - h_1^{-2} \delta_j \triangle P_j^{K-1,K} E\left[\rho_{g\delta}(m_i) \ddot{H}_{1h_1}(m_{ji}) | X_j\right] \\ &= h_1^{-2} \delta_j \int \ddot{H}_1(u) \rho(m_j + hu) f_m(m_j + hu) \, du \\ &\quad - h_1^{-2} \delta_j \triangle P_j^{K-1,K} \int \ddot{H}_1(u) \rho_{g\delta}(m_j + hu) f_m(m_j + hu) \, du \end{aligned}$$

$$= \frac{1}{2}\delta_j \partial^2 \left[\rho(m_j) f_m(m_j)\right] + \frac{1}{2}\delta_j \triangle P_j^{K-1,K} \partial^2 \left[\rho_{g\delta}(m_j) f_m(m_j)\right] + r_{3n1,a}(X_j, \pi),$$

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 $\int \ddot{H}_1(u)\, u du = 0$. With the above results, we can readily show that

$$\begin{aligned} |\theta_{3n}(\pi)| &= |E[r_{3n,1}(X_j,\pi) + r_{3n,2}(X_j,\pi)]| \\ &\lesssim \frac{1}{2}\left|E\left\{\delta_j \triangle g_j \partial^2\left[\rho_{\delta P}(m_j) f_m(m_j)\right]\right\}\right| + \frac{1}{2}\left|E\left\{\delta_j \triangle g_j \triangle P_j^{K-1,K} \partial^2\left[\rho_\delta(m_j) f_m(m_j)\right]\right\}\right| \\ &\equiv \theta_{3n,1}(\pi) + \theta_{3n,2}(\pi). \end{aligned}$$

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

$$\begin{aligned} & E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \|q_{3n}((X_1,X_2),\pi)\|^2\right] \\ \lesssim\ & E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\triangle g_{1,2} \triangle P_{1,2}^{K-1,K} h_1^{-2} \ddot{H}_{1h_1}(m_{21}) \left[L\left(R(X_1)'\pi\right) - m_1\right] \left[L\left(R(X_2)'\pi\right) - m_2\right]\right\|^2\right] \end{aligned}$$

$$\lesssim \max_{1\leq i\leq n}\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left\|L\left(R(X_i)'\pi\right)-m_i\right\|^4 E\left[\left\|\triangle g_{1,2}\triangle P_{1,2}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}\left(m_{21}\right)\right\|^2\right]$$
$$= O\left(K^2\eta_{1Kn}^4\right)O\left(Kh^{-5}\right)=O\left(h_1^{-5}K^3\eta_{1Kn}^4\right).$$

Then by Corollary 5.3 in Chen and Kato (2020), we have

$$\sup_{\|\pi-\pi^0\|\leq C\eta_{Kn}}\left\|\mathbb{U}_{3n}^{(2)}\left(\pi\right)\right\|\lesssim n^{-1}\left({h_1}^{-5}K^3\eta_{1Kn}^4\right)^{1/2}\ln\left(n\right)=o_p\left(\eta_{1Kn}\right).$$

By the empirical process theory, we have

$$\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left\|\mathbb{U}_{3n}^{(1)}\left(\pi\right)\right\|=O_p\left(n^{-1/2}K^{1/2}\ln\left(n\right)\right)O\left(K^{1/2}\eta_{1Kn}\right)=o_p\left(\eta_{1Kn}\right).$$

Consequently, we have

$$\left\|U_{3n}\left(\bar{\pi}\right)\right\|\lesssim\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left\|\theta_{3n}\left(\pi\right)+2\mathbb{U}_{3n}^{(1)}\left(\pi\right)+\mathbb{U}_{3n}^{(2)}\left(\pi\right)\right\|=o_p\left(\eta_{1Kn}\right)$$

and $\|L_{1,n4}\|=o_p\left(\eta_{1Kn}\right)$. Analogously, we can show that $\|L_{1,n\ell}\|=o_p\left(\eta_{1Kn}\right)$ for $\ell=5,6,7$.

(iv) The result follows from (i)-(iii). ■

Lemma B.3 *Let $L_{2,n}$, $L_{2,n1},...,L_{2,n7}$ be as defined in the proof of Theorem 3.1. Suppose that the conditions in Theorem 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

$$\begin{aligned}\|L_{2,n1}\| &= \left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}H_{1h_1,ji}\right\| \\ &\leq 2\max_i\left|\triangle P_i^{K\prime}\beta_0-\triangle g_i\right|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left\|\triangle P_{i,j}^{K-1,K}\right\|\left|H_{1h_1,ji}\right| \\ &= O_p\left(K^{-\gamma}\right)O_p\left(K^{1/2}\right)=O_p\left(K^{-\gamma+1/2}\right)\end{aligned}$$

(ii) Note that

$$\begin{aligned}\|L_{2,n2}\| &= \left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}h_1^{-1}\dot{H}_{1h_1,ji}\left(\bar{m}_j-m_j\right)\right\| \\ &\leq 2\max_i\left|\triangle P_i^{K\prime}\beta_0-\triangle g_i\right|\max_j\left|\bar{m}_j-m_j\right|\frac{h_1^{-1}}{n^2}\sum_{1\leq i\neq j\leq n}\left\|\triangle P_{i,j}^{K-1,K}\right\|\left|\dot{H}_{1h_1,ji}\right| \\ &= O_p\left(K^{-\gamma}\right)O_p\left(K^{1/2}\eta_{1Kn}\right)O_p\left(h_1^{-1}\right)=o_p\left(K^{-\gamma}\right).\end{aligned}$$

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.4 *Let $G_{1,nj},...,G_{4,nj}$ be as defined in (A.19) in the proof of Theorem 3.2. Suppose that the conditions in Theorem 3.2 are satisfied. Then*

(i) There exists a positive constant c_F such that $\|G_{1,nj}\| \leq c_F |\triangle g_j - \triangle\bar{g}_j| + O_p(h_2^{a_2+1})$ uniformly in $j = 1, 2, ..., n$;

(ii) There exists a positive constant c_F such that $\|G_{2,nj}\| \leq c_F |\triangle g_j - \triangle\bar{g}_j| + O_p(h_2^{a_2+1})$ uniformly in $j = 1, 2, ..., n$;

(iii) $\max_{1\leq j\leq n} \|G_{3,nj}\| \leq O_p(\eta_{2Kn})$;

(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

$$\begin{aligned}
&\|G_{1,nj}\| \\
&\leq \Bigg|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)' \\
&\times L'(LF(\triangle\bar{g}_i))\left\{\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\left[\partial^l LF(\triangle g_j) - \partial^l LF(\triangle\bar{g}_j)\right]/l! - \sum_{l=a_2+1}^{\infty}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l \partial^l LF(\triangle\bar{g}_j)/l!\right\}\Bigg| \\
&\leq \sum_{l=0}^{a_2}\left|\partial^l LF(\triangle g_j) - \partial^l LF(\triangle\bar{g}_j)\right|/l! \\
&\times \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)' L'(LF(\triangle\bar{g}_i))\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l/l!\right\| \\
&+ \left\|\sum_{l=a_2+1}^{\infty}\left|\partial^l LF(\triangle\bar{g}_j)\right|/l!\left|\frac{1}{n}\sum_{i=1}^{N} H_{h_2}(\triangle\bar{g}_{i,j})\left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)' L'(LF(\triangle\bar{g}_i))\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\right|\right\| \\
&\equiv G_{1,nj,1} + G_{1,nj,2}.
\end{aligned}$$

By the uniform boundedness of all finite order derivatives of $L(\cdot)$,

$$\begin{aligned}
\|G_{1,nj,1}\| &\leq C|\triangle g_j - \triangle\bar{g}_j|\sum_{l=0}^{a_2}\left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)\left(\frac{\triangle\bar{g}_{i,j}}{h_2}\right)^l\right\| \\
&\leq C|\triangle g_j - \triangle\bar{g}_j|\sum_{l=0}^{a_2}\frac{1}{n}\sum_{i=1}^{N}\|H_{2h_2}(\triangle\bar{g}_{i,j})\| \leq C|\triangle g_j - \triangle\bar{g}_j|
\end{aligned}$$

where recall that C can vary over places. For $G_{1,nj,2}$, we have

$$\|G_{1,nj,2}\| \lesssim \left\|\sum_{l=a_2+1}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{N}|H_{h_2}(\triangle\bar{g}_{i,j})|\left\|\left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)\right\|\left(\frac{\triangle\bar{g}_{i,j}}{h_2}\right)^l\right|\right\|$$

$$= O_p(h_2^{a_2+1})$$

where we use the fact that $\max_i |\triangle\bar{g}_i| = O_p\left(h_2^2\right)$. It follows that $G_{1,nj} \leq c_F |\triangle g_j - \triangle\bar{g}_j| + O_p(h_2^{a_2+1})$ 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 3.1(iv),

$$\begin{aligned} |G_{2,nj,02}| &\lesssim h_2^{-1} \max_{i,j} |\triangle\bar{g}_{i,j} - \triangle g_{i,j}| \max_i |\triangle\bar{g}_i - \triangle g_i| \\ &= O_p\left(h_2^{-1}\left(K^{1/2}\eta_{2Kn}\right)^2\right) = O_p\left(\eta_{2Kn}\right) \text{ uniformly in } j. \end{aligned}$$

For $G_{2,nj,01}$, we can show that

$$\begin{aligned} |G_{2,nj,01}| &\lesssim \left|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle g_{i,j}\right)\left\{F\left(\triangle\bar{g}_i\right) - F\left(\triangle g_i\right)\right\}\right| \\ &\lesssim \left|\frac{1}{n}\sum_{i=1}^{N} \left|H_{2h_2}\left(\triangle g_{i,j}\right) F'\left(\triangle g_i\right)\right|\left(\triangle\bar{g}_i - \triangle g_i\right)\right| = O_p\left(\eta_{2Kn}\right). \end{aligned}$$

Then $|G_{2,nj,0}| = O_p\left(\eta_{2Kn}\right)$. Similarly, we can show that $|G_{2,nj,0\ell}| = O_p\left(\eta_{2Kn}\right)$ for $\ell = 1, ..., a_2$. It follows that $|G_{2,nj}| = O_p\left(\eta_{2Kn}\right)$.

(iii) The proof is analogous to that of (i) and thus omitted.

(iv) Recall that $\triangle\bar{g}_i = \bar{g}(X_{i2}) - \bar{g}\left(X_{i1}\right)$ and $\triangle\bar{g}_{i,j} = \triangle\bar{g}_i - \triangle\bar{g}_j$. Let $\triangle g_i = g(X_{i2}) - g\left(X_{i1}\right)$ and $\triangle g_{i,j} = \triangle g_i - \triangle g_j$. Note that $G_{4,nj} = \left(G_{4,nj,0}, ..., G_{4,na_S}\right)$, where

$$G_{4,nj,l}^0 = \frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left(\frac{1}{h_2}\triangle\bar{g}_{i,j}\right)^l \left[F(\triangle g_i) - D_i\right].$$

Let $G_{4,nj,l}^0 = \frac{1}{n}\sum_{i=1, i\neq j}^{N} H_{h_2}\left(\triangle g_{i,j}\right)\left(\frac{1}{h_2}\triangle g_{i,j}\right)^l [F(\triangle g_i) - D_i]$ for $l = 0, 1, ..., a_2$ where 0^0 is defined to be 1. Noting that $E\left(D_i|X_i\right) = F\left(\triangle g_i\right)$, we can apply Bernstein exponential inequality to show that

$$\max_{1\leq j\leq n} \left\|G_{4,nj,l}^0\right\| = O_p\left(\sqrt{\log(n)/(nh_2)}\right) \text{ for } l = 0, 1, ..., a_2.$$

Next,

$$G_{4,nj,0}-G^0_{4,nj,0}=\frac{1}{n}\sum_{i=1,i\neq j}^{N}\left[H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)-H_{2h_2}\left(\triangle g_{i,j}\right)\right]\left[F(\triangle g_i)-D_i\right]-\frac{1}{n}H_{2h_2}\left(0\right)\left[F(\triangle g_j)-D_j\right].$$

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 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 and show it is $o_p\left(\sqrt{\log(n)/(nh_2)}\right)$ uniformly in j. ■

References

Bach, F., 2010. Self-concordant analysis for logistic regression. Electronic Journal of Statistics 4, 384–414.

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.

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.

Chen, X., Kato, K., 2020. Jackknife multiplier bootstrap: finite sample approximations to the u-process supremum with applications. Probability Theory and Related Fields 176, 1097–1163.

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.

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., 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.

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

Jiang, L., Phillips, P.C., Tao, Y., Zhang, Y., 2021. Regression-adjusted estimation of quantile treatment effects under covariate-adaptive randomizations. arXiv preprint arXiv:2105.14752 .

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.

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.

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