cemmap

centre for microdata methods and practice

Estimation in semiparametric quantile factor models

Shujie Ma
Oliver Linton

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP07/18

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Estimation in Semiparametric Quantile Factor Models

Shujie Ma

University of California at Riverside

Oliver Linton

University of Cambridge

Jiti Gao

Monash University*

May 5, 2017

Abstract

We propose an estimation methodology for a semiparametric quantile factor panel model. We provide tools for inference that are robust to the existence of moments and to the form of weak cross-sectional dependence in the idiosyncratic error term. We apply our method to CRSP daily data.

Keywords: Cross–Sectional Dependence; Fama–French Model; Inference; Sieve Estimation

*The research of the third author was supported by the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421.

1 Introduction

In a series of papers, Fama and French (1992,1993,1995,1996,1998) developed a general methodology for estimating factor panel models for stock returns and for testing the Arbitrage Pricing Theory, which has been extremely influential. Connor and Linton (2007) and Connor, Hagmann and Linton (2012) developed a semiparametric panel regression methodology to describe the same phenomenon, but with the feature that stock characteristics were used explicitly inside a model, which then allowed proper inferential procedures that account fully for the sampling uncertainty. Specifically, they introduced a semiparametric characteristic-based factor model in which the factor betas are smooth functions of a small number of observable characteristics, while the factor returns are estimable quantities. Their estimation methodology is based on two steps: estimating the beta functions using nonparametric kernel smoothing for additive regression given the factor returns, and second, estimating the factor returns by OLS or GLS given the estimated beta functions. They established some large sample properties of their procedure and applied it to the same monthly data used in FF, finding improved results. In addition, because their work was based on an explicit regression model, they were able to give standard errors that accounted correctly for the sampling variability in their estimates. This methodology was based on least squares concepts and made use of projection arguments. They required at least four moments to establish their CLT, which may not be a binding restriction for monthly stock returns. However, for daily stock returns this is be a bit strong, especially for small caps.

In the empirical literature, there is a lot of interest in applying factor models to daily data. Perhaps the current state of the art for factor modelling proposed by Fan, Lv, and Mikusheva (2013) extended the work of Bai and Ng (2002) by allowing the idiosyncratic covariance matrix to be non-diagonal but sparse, and used thresholding techniques to impose sparsity and thereby obtain a better estimator of the covariance matrix and its inverse in this big-data setting. They also imposed many moments on the return series for their theoretical analysis, although they applied their techniques to daily data. Quantile methods are widely used in economics and finance, see, for example, Koenker and Bassett (1978); indeed, they are classified as "harmless econometrics", see Angrist and Pischke (2009). They have the advantage of being robust to large observations. Boneva, Linton, and Vogt (2015) have applied quantile techniques to a linear in parameters panel model with unobserved effects, extending Pesaran (2006). Sharma, Gupta, and Singh (2016) applied linear quantile regression to estimate a four factor FF "model" to daily Indian data from 1993-2016. They found that not all factors are substantially present across all quantiles, which adds some colour to the usual mean results. Horowitz and Lee (2005) defined an estimaiton method for additive quantile regression. Belloni, Chernozhukov and Fernandez-Val (2016) have recently proposed a number of inference methods for quantile regression with a nonparametric component or a large number of unknown parameters, but their tools are developed within a cross-sectional iid setting and so do not directly apply here.

In this paper, we propose estimation and inferential methodology for the quantile version of the Connor and Linton (2007) semiparametric panel model for financial returns, which does not require such strong moment restrictions, thereby facilitating work with daily data. Our contribution is summarized as follows.

First, we propose an estimation algorithm for this model. We use sieve techniques to obtain preliminary estimators of the nonparametric beta functions, see Chen (2011) for a review, and then update each component sequentially. We compute the estimator in two steps for computational reasons. We have $J \times T$ unknown factor return parameters as well as $J \times K_N$ sieve parameters to estimate, and to estimate these simultaneously without penalization would be challenging. Penalization of the factor returns here would not be well motivated so we do not pursue this. Instead we first estimate the unrestricted additive quantile regression function for each time period and then impose the factor structure in a sequential fashion.

Second, we derive the limiting properties of our estimated factor returns and factor loading functions under the assumption that the included factors all have non zero mean and under weak conditions on cross-section and temporal dependence. A key consideration in the panel modelling of stock returns is what position to take on the cross sectional dependence in the idiosyncratic part of stock returns. Early studies assumed iid in the cross section, but this turns out not to be necessary. More recent work has allowed for cross sectional dependence in a variety of ways. Connor, Hagmann and Linton (2012) imposed a known industry cluster/block structure where the number of industries goes to infinity as do the number of members of the industry. Under this structure one obtains a CLT and inference can be conducted by estimating only the intra block covariances. Robinson and Thawornkaiwong (2012) considered a linear process structure driven by independent shocks. Dong, Gao and Peng (2015) introduced a spatial mixing structure to accommodate both serial correlation and cross–sectional dependence for a general panel data setting. Under a lattice structure or some observable or estimable distance function that determines the ordering, Conley (1999), one can consistently estimate the asymptotic covariance matrix. However, this type of structure is hard to justify for stock returns, and in that case their approach does not deliver consistent inference. Connor and Koraczyck (1993) considered a different cross-sectional dependence structure, namely they supposed that there was an ordering of the cross sectional units such that weak dependence of the alpha mixing variety was held. They do not assume knowledge of the ordering as this was not needed for their main results. We adopt and generalize their structure. In fact, we allow for weak dependence simultaneously in the cross-section and time series dependence. This structure affects the limiting distribution of the estimated factor returns in a complicated fashion, and the usual Newey–West type of standard errors can't be adapted to account for the cross-sectional dependence here because the ordering is not assumed to be known. To conduct inference we have to take account of the correlation structure. We use the so-called fix-b asymptotics to do this, namely we construct a test statistic based on an inconsistent fixed-b kitchen sink estimator of the correlation structure, as in Kiefer and Vogelsang (2002), and show that it has a pivotal limiting distribution that is a functional of a Gaussian process.

Third, our estimation procedure requires only that the time series mean of factor returns be non zero. A number of authors have noted that in the presence of a weak factor, regression identification strategies can break down, Bryzgalova (2015). In view of this we provide a test of whether a given factor is present or not in each time period. Fourth, we apply our procedure to CRSP daily data and show how the factor loading functions vary nonlinearly with state. The median regression estimators are comparable to those of Connor, Hagmann and Linton (2012) and can be used to test asset pricing theories under comparable quantile restrictions, see for example, Bassett, Koenker and Kordas (2004), and to design investment strategies. The lower quantile estimators could be used for risk management purposes. The advantage of the quantile method is its robustness to heavy tails in the response distribution, which may be present in daily data. Indeed our theory does not require any moment conditions.

The organization of this paper is given as follows. Section 2 proposes the main model and then discusses some identification issues. An estimation method based on B–splines is then proposed in Section 3. Section 4 establishes an asymptotic theory for the proposed estimation method. Section 5 discusses a covariance estimation problem and then considers testing for the factors involved in the main model. Section 6 gives an empirical application of the proposed model and estimation theory to model the dependence of daily returns on a set of characteristic variables. Section 7 concludes the paper with some discussion. All the mathematical proofs of the main results are given in an appendix.

2 The model and identification

We introduce some notations which will be used throughout the paper. For any positive numbers a_n and b_n, let $a_n \asymp b_n$ denote $\lim_{n\to\infty} a_n/b_n = c$, for a positive constant c, and let $a_n \gg b_n$ denote $a_n^{-1} b_n = o(1)$. For any vector $\mathbf{a} = (a_1, \ldots, a_n)^\intercal \in \mathbb{R}^n$, denote $||\mathbf{a}|| = (\sum_{i=1}^n a_i^2)^{1/2}$. For any symmetric matrix $\mathbf{A}_{s\times s}$, denote its L_2 norm as $\|\mathbf{A}\| = \max_{\zeta\in\mathbb{R}^s,\zeta\neq\mathbf{0}} \|\mathbf{A}\zeta\| \, \|\zeta\|^{-1}$. We use

$(N, T) \to \infty$ to denote that N and T pass to infinity jointly.

We consider the following model for the τ^{th} conditional quantile function of the response y_{it} for the i^{th} asset at time t given as

$$Q_{y_{it}}(\tau|X_i, f_t) = f_{ut}^0 + \sum\nolimits_{j=1}^{J} g_j^0(X_{ji}) f_{jt}^0, \tag{2.1}$$

i.e., we suppose that

$$y_{it} = f_{ut} + \sum\nolimits_{j=1}^{J} g_j(X_{ji}) f_{jt} + \varepsilon_{it}, \tag{2.2}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where y_{it} is the excess return to security i at time t; f_{ut} and f_{jt} are factor returns, which are unobservable; $g_j(X_{ji})$ are the factor betas, which are unknown but smooth functions of X_{ji}, where X_{ji} are observable security characteristics, and X_{ji} lies in a compact set $\mathcal{X}_{ji}$. The error terms ε_{it} are the asset-specific or idiosyncratic returns and they satisfy that the conditional τ^{th} quantile of ε_{it} in (2.2) given (X_i, f_t) is zero. The factors f_{ut}^0 and f_{jt}^0 and the factor betas $g_j^0(\cdot)$ should be τ specific. For notational simplicity, we suppress the τ subscripts. For model identifiability, we assume that:

Assumption A0. *For some probability measures P_j we have $\int g_j^0(x_j) dP_j(x_j) = 0$ and* $\int \left(g_j^0(x_j)\right)^2 dP_j(x_j) = 1$ for all $j = 1, \ldots, J$. Furthermore, $\liminf_{T\to\infty} \left|\sum_{t=1}^{T} f_{jt}^0/T\right| > 0$ for each j.

The case where $\tau = 1/2$ corresponds to the conditional median, and is broadly comparable to the conditional mean model used in Connor and Linton (2007) and Connor, Hagmann and Linton (2012). The advantage of the median over the mean is its robustness to heavy tails and outliers, which is especially important with daily data. The case where $\tau = 0.01$, say, might be of interest for the purposes of risk management, since this corresponds to a standard Value-at-Risk threshold in which case (2.1) gives the conditional Value-at-Risk given the characteristics and the factor returns at time t. To obtain an ex-ante measure we should have to employ a forecasting model for the factor returns.

Suppose that the τ^{th} conditional quantile function $Q_{y_{it}}(\tau|X_i = x)$ of the response y_{it} at time t given the covariate $X_i = x$ is additive

$$H_t(\tau|x) = h_{ut}^0 + \sum\nolimits_{j=1}^{J} h_{jt}^0(x_j), \tag{2.3}$$

where $h_{jt}^0(\cdot)$ are unknown functions without loss of generality satisfying $\int h_{jt}^0(x_j) dP_j(x_j) = 0$ for $t = 1, \ldots, T$, Horowitz and Lee (2005). Under the factor structure (2.1), we have for all j

$$\int \left(\frac{1}{T}\sum_{t=1}^{T} h_{jt}^0(x_j)\right)^2 dP_j(x_j) = \int g_j^0(x_j)^2 dP_j(x_j) \times \left(\frac{1}{T}\sum_{t=1}^{T} f_{jt}^0\right)^2 = \left(\frac{1}{T}\sum_{t=1}^{T} f_{jt}^0\right)^2. \tag{2.4}$$

Provided $\sum_{t=1}^{T} f_{jt}^0 \neq 0$, then we may identify $g_j^0(x_j)$ by

$$g_j^0(x_j) = \frac{\frac{1}{T}\sum_{t=1}^{T} h_{jt}^0(x_j)}{\sqrt{\int \left(\frac{1}{T}\sum_{t=1}^{T} h_{jt}^0(x_j)\right)^2 dP_j(x_j)}}. \tag{2.5}$$

We will use this as the basis for estimation.

Note that the identification strategy also works in the case where X_i also varies over t.

3 Estimation

3.1 Factor returns and characteristic-beta functions

We propose an iterative algorithm to estimate the factor returns and the characteristic-beta function. The algorithm makes use of the structure so as to minimize the dimensonality of the optimization problems involved. The right hand side of (2.1) is bilinear in unknown quantities, and so it seems hard to avoid such an algorithmic approach.

To estimate $g_j^0(\cdot)$, we first approximate them by B-spline functions described as follows. Let $b_j(x_j) = \{b_{j,1}(x_j), \ldots, b_{j,K_N}(x_j)\}^\intercal$ be a set of normalized B-spline functions of order m (see, for example, de Boor (2001)), where $K_N = L_N + m$, and L_N is the number of interior knots satisfying $L_N \to \infty$ as $N \to \infty$. We adopt the centered B-spline basis functions $B_j(x_j) = \{B_{j,1}(x_j), \ldots, B_{j,K_N}(x_j)\}^\intercal$, where

$$B_{jk}(x_j) = \sqrt{K_N}\left[b_{j,k}(x_j) - N^{-1}\sum\nolimits_{i=1}^{N} b_{j,k}(X_{ji})\right],$$

so that $N^{-1}\sum_{i=1}^{N} B_{jk}(X_{ji}) = 0$ and $\mathrm{var}\{B_{jk}(X_j)\} \asymp 1$. We first approximate the unknown functions $g_j(x_j)$ by B-splines such that $g_j(x_j) \approx B_j(x_j)^\intercal \boldsymbol{\lambda}_j$, where $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \ldots, \lambda_{j,K_N})^\intercal$ are spline coefficients. Hence $N^{-1}\sum_{i=1}^{N} B_j(X_{ji})^\intercal \boldsymbol{\lambda}_j = 0$. Denote $f_t = \{f_{ut}, (f_{jt}, 1 \le j \le J)^\intercal\}^\intercal$. Let $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\intercal, \ldots, \boldsymbol{\lambda}_J^\intercal)^\intercal$ and let $\rho_\tau(u) = u(\tau - I(u < 0))$ be the quantile check function. The iterative algorithm is described as follows:

1. Find the initial estimates $\widehat{f}^{[0]}$ and $\widehat{g}_j^{[0]}(\cdot)$.
2. For given $\widehat{f}^{[i]}$, we obtain

$$\widehat{\boldsymbol{\lambda}}^{[i+1]} = \arg\min_{\boldsymbol{\lambda}\in\mathbb{R}^{JK_N}} \sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} \rho_\tau\left(y_{it} - \widehat{f}_{ut}^{[i]} - \sum\nolimits_{j=1}^{J} B_j(X_{ji})^\intercal \boldsymbol{\lambda}_j \widehat{f}_{jt}^{[i]}\right).$$

Let $\widehat{g}_j^{*[i+1]}(x_j) = B_j(x_j)^\intercal \widehat{\boldsymbol{\lambda}}_j^{[i+1]}$. The estimate for $g_j(x_j)$ at the $(i+1)^{\text{th}}$ step is

$$\widehat{g}_j^{[i+1]}(x_j) = \frac{\widehat{g}_j^{*[i+1]}(x_j)}{\sqrt{N^{-1}\sum_{i=1}^{N} \widehat{g}_j^{*[i+1]}(X_{ji})^2}}.$$

3. For given $\widehat{g}_j^{[i+1]}(x_j)$, we obtain for $t = 1, \ldots, T$

$$\widehat{f}_t^{[i+1]} = \arg\min_{f_t \in \mathbb{R}^{J+1}} \sum_{i=1}^{N} \rho_\tau \left(y_{it} - f_{ut} - \sum_{j=1}^{J} \widehat{g}_j^{[i+1]}(X_{ji}) f_{jt} \right).$$

We repeat steps 2 and 3, and consider that the algorithm converges at the $(i+1)^{\text{th}}$ step when $||\widehat{f}^{[i+1]} - \widehat{f}^{[i]}|| < \epsilon$ and $||\widehat{\boldsymbol{\lambda}}^{[i+1]} - \widehat{\boldsymbol{\lambda}}^{[i]}|| < \epsilon$ for a small positive value ϵ. Then the final estimates are $\widehat{f}_t = \widehat{f}_t^{[i+1]}$ and $\widehat{g}_j(x_j) = \widehat{g}_j^{[i+1]}(x_j)$. Our experience in numerical analysis suggests that the proposed method converges well and rapidly. Below, we propose a way to obtain consistent initial starting values. The algorithm can stop after a finite number of iterations by using the consistent initial values.

3.2 Initial estimators

Let $b_j(x_j) = \{b_{j,1}(x_j), \ldots, b_{j,R_N}(x_j)\}^\intercal$ be a set of normalized B-spline functions of order m (22), where R_N is the number of interior knots satisfying $R_N \to \infty$ as $N \to \infty$. We adopt the centered B-spline basis functions $B_j(x_j) = \{B_{j,1}(x_j), \ldots, B_{j,R_N}(x_j)\}^\intercal$, where

$$B_{jk}(x_j) = \sqrt{R_N} \left[b_{j,k}(x_j) - N^{-1} \sum_{i=1}^{N} b_{j,k}(X_{ji}) \right],$$

so that $N^{-1} \sum_{i=1}^{N} B_{jk}(X_{ji}) = 0$ and $\text{var}\{B_{jk}(X_j)\} \asymp 1$. We first approximate the unknown functions $h_{jt}(x_j)$ by B-splines such that $h_{jt}(x_j) \approx B_j(x_j)^\intercal \boldsymbol{\vartheta}_j$, where $\boldsymbol{\vartheta}_j = (\vartheta_{j,1}, \ldots, \vartheta_{j,R_N})^\intercal$ are spline coefficients. Hence $N^{-1} \sum_{i=1}^{N} B_j(X_{ji})^\intercal \boldsymbol{\vartheta}_j = 0$. Let $\rho_\tau(u) = u(\tau - I(u < 0))$ be the quantile check function. Let

$$\widehat{\boldsymbol{\vartheta}}_t = (\widehat{\vartheta}_{0,t}, \widehat{\boldsymbol{\vartheta}}_{1,t}^\intercal, \ldots, \widehat{\boldsymbol{\vartheta}}_{J,t}^\intercal)^\intercal = \arg\min_{\boldsymbol{\vartheta} = (\vartheta_0, \boldsymbol{\vartheta}_1^\intercal, \ldots, \boldsymbol{\vartheta}_J^\intercal)^\intercal \in \mathbb{R}^{JR_N+1}} \sum_{i=1}^{N} \rho_\tau \left(y_{it} - \vartheta_0 - \sum_{j=1}^{J} B_j(X_{ji})^\intercal \boldsymbol{\vartheta}_j \right),$$

and let $\widehat{h}_{jt}(x_j) = B_j(x_j)^\intercal \widehat{\boldsymbol{\vartheta}}_{j,t}$.

Let $\widetilde{h}_{ut}$ and $\widetilde{h}_{jt}(X_{ji})$ be the estimators of h_{ut}^0 and $h_{jt}^0(X_{ji})$ from fitting the quantile regression model (2.3). We let the initial estimators of $g_j^0(x_j)$ be

$$\widehat{g}_j^{[0]}(x_j) = \frac{\frac{1}{T} \sum_{t=1}^{T} \widetilde{h}_{jt}(x_j)}{\sqrt{\int \left(\frac{1}{T} \sum_{t=1}^{T} \widetilde{h}_{jt}(x_j) \right)^2 dP_j(x_j)}}. \tag{3.1}$$

We use the spline smoothing method to obtain the estimators $\widetilde{h}_{jt}(x_j)$.We first approximate the unknown functions $h_{jt}(x_j)$ by B-splines such that $h_{jt}(x_j) \approx B_j(x_j)^\intercal \boldsymbol{\theta}_{jt}$, where $\boldsymbol{\theta}_{jt} = (\theta_{jt,1}, \ldots, \theta_{jt,K_N})^\intercal$ are spline coefficients. Let $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{1t}^\intercal, \ldots, \boldsymbol{\theta}_{Jt}^\intercal)^\intercal$. Then the estimators $(\widetilde{h}_{ut}, \widetilde{\boldsymbol{\theta}}_t^\intercal)^\intercal$ of $(h_{ut}, \boldsymbol{\theta}_t^\intercal)^\intercal$ are obtained by minimizing

$$\sum_{i=1}^{N} \rho_\tau (y_{it} - h_{ut} - \sum_{j=1}^{J} B_j(X_{ji})^\intercal \boldsymbol{\theta}_{jt}) \tag{3.2}$$

with respect to $(h_{ut}, \boldsymbol{\theta}_t^\intercal)^\intercal \in \mathbb{R}^{JK_N}$. Then the estimator of $h_{jt}^0(x_j)$ is $\widetilde{h}_{jt}(x_j) = B_j(x_j)^\intercal \widetilde{\boldsymbol{\theta}}_{jt}$. The initial estimator for f_t is

$$\widehat{f}_t^{[0]} = \arg\min_{f_t \in \mathbb{R}^{J+1}} \sum_{i=1}^N \rho_\tau(y_{it} - f_{ut} - \sum_{j=1}^J \widehat{g}_j^{[0]}(X_{ji}) f_{jt}) \tag{3.3}$$

for $t = 1, \ldots, T$.

4 Asymptotic theory of the estimators

We suppose that there is some relabelling of the cross-sectional units $i_{l_1}, \ldots, i_{l_N}$, whose generic index we denote by i^*, such that the cross sectional dependence decays with the distance $|i^* - j^*|$. This assumption has been made in Connor and Korajczyk (1993). There are available algorithms to determine the true ordering from the original ordering given sample data (and under the assumption that this ordering is monotonic), but we shall not pursue this, because it will not be necessary for estimation or inference to know this ordering. In fact we will allow dependence both across time and in the cross-section. For notational simplicity, we denote the indices as $\{i, 1 \le i \le N\}$ after the ordering. Let $\mathbb{N}$ denotes the collection of all positive integers. We use a ϕ-mixing coefficient to specify the dependence structure. Let $\{W_{it} : 1 \le i \le N, 1 \le t \le T\}$, where $W_{it} = (X_i^\intercal, f_t^\intercal, \varepsilon_{it})^\intercal$ and $\varepsilon_{it} = y_{it} - f_{ut}^0 - \sum_{j=1}^J g_j^0(X_{ji}) f_{jt}^0$. For $S_1, S_2 \subset [1, \ldots, N] \times [1, \ldots, T]$, let

$$\phi(S_1, S_2) \equiv \sup\{|P(A|B) - P(A)| : \\ A \in \sigma(W_{it}, (i,t) \in S_1), B \in \sigma(W_{it}, (i,t) \in S_2)\},$$

where $\sigma(\cdot)$ denotes a σ-field. Then the ϕ-mixing coefficient of $\{W_{it}\}$ for any $k \in \mathbb{N}$ is defined as

$$\phi(k) \equiv \sup\{\phi(S_1, S_2) : d(S_1, S_2) \ge k\},$$

and

$$d(S_1, S_2) \equiv \min\{\sqrt{|t-s|^2 + |i-j|^2} : (i,t) \in S_1, (j,s) \in S_2\}.$$

Without loss of generality, we assume that $\mathcal{X}_{ji} = [a, b]$. Denote $h_t^0(x) = \{h_{jt}^0(x_j), 1 \le j \le J\}^\intercal$ and $\widetilde{h}_t(x) = \{\widetilde{h}_{jt}(x_j), 1 \le j \le J\}^\intercal$, where $x = (x_1, \ldots, x_J)^\intercal$. Let $G_i^0(X_i) = \{1, g_1^0(X_{1i}), \ldots, g_J^0(X_{Ji})\}^\intercal$. We make the following assumptions.

(C1) $\{W_{it}\}$ is a random field of ϕ-mixing random variables. The ϕ-mixing coefficient of $\{W_{it}\}$ satisfies $\phi(k) \le K_1 e^{-\lambda_1 k}$ for $K_1, \lambda_1 > 0$. For each given i, $\{W_{it}\}$is a strictly stationary sequence.

(C2) The conditional density $p_i(\varepsilon \,|x_i, f_t)$ of ε_{it} given (x_i, f_t) satisfies the Lipschitz condition of order 1 and $\inf_{1 \le i \le N, 1 \le t \le T} p_i(0 \,|x_i, f_t) > 0$. For every $1 \le j \le J$, the density function $p_{X_{ji}}(\cdot)$ of X_{ji} is bounded away from 0 and satisfies the Lipschitz condition of order 1 on $[a,b]$. The density function $f_{X_i}(\cdot)$ of X_i is absolutely continuous on $[a,b]^J$.

(C3) The functions g_j^0 and h_{jt}^0 are r-times continuously differentiable on its support for some $r > 2$. The spline order satisfies $m \geq r$.

(C4) There exist some constants $0 < c_h \leq C_h < \infty$ and $0 < C_h' < \infty$ such that $c_h \leq \left(\frac{1}{T}\sum_{t=1}^{T} f_{jt}^0\right)^2 \leq C_h$ for all j with probability tending to one.

(C5) The eigenvalues of the $(J+1)\times(J+1)$ matrix $N^{-1}\sum_{i=1}^{N} E(G_i^0(X_i)G_i^0(X_i)^\intercal)$ are bounded away from zero almost surely.

(C6) Let Ω_N^0 be the covariance matrix of $N^{-1/2}\sum_{i=1}^{N} G_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))$. The eigenvalues of Ω_N^0 are bounded away from zero and infinity almost surely.

We allow that $\{W_{it}\}$ are weakly dependent across i and t, but need to satisfy the strong mixing condition given in Condition (C1). Moreover, Condition (C1) implies that $\{X_i\}$ is marginally cross-sectional mixing, and $\{f_t\}$ is marginally temporally mixing. Similar assumptions are used in Gao, Lu and Tjøstheim (2006) for an alpha–mixing condition in a spatial data setting, and Dong, Gao and Peng (2016) for introducing a spatial mixing condition in a panel data setting. Conditions (C2) and (C3) are commonly used in the nonparametric smoothing literature, see for example, Horowitz and Lee (2005), and Ma, Song and Wang (2013). Condition (C4) and (C5) are similar to Conditions A2, A5 and A7 of Connor, Matthias and Linton (2012).

Let 1_l be the $(J+1)\times 1$ vector with the l^{th} element as "1" and other elements as "0". Denote $B(X_i) = \{B_1(X_{1i})^\intercal, \ldots, B_J(X_{Ji})^\intercal\}^\intercal$ and

$$Z_i = [\{1, B(X_i)^\intercal\}^\intercal]_{(1+JK_N)\times 1}. \tag{4.1}$$

Let

$$\mathbb{B}(x) = [\text{diag}\{1, B_1(x_1)^\intercal, \ldots, B_J(x_J)^\intercal\}]_{(1+J)\times(1+JK_N)}. \tag{4.2}$$

Define

$$\Lambda_{Nt}^0 = N^{-1}\sum_{i=1}^{N} E\{p_i\,(0\,|X_i, f_t)\, G_i^0(X_i)G_i^0(X_i)^\intercal\}. \tag{4.3}$$

and

$$\mathbf{\Sigma}_{Nt}^0 = \tau(1-\tau)(\Lambda_{Nt}^0)^{-1}\Omega_N^0(\Lambda_{Nt}^0)^{-1}. \tag{4.4}$$

The theorem below presents the asymptotic distribution of the final estimator $\widehat{f}_t$. Define

$$\phi_{NT} = \sqrt{K_N/(NT)} + K_N^{3/2}N^{-3/4}\sqrt{\log NT} + K_N^{-r}. \tag{4.5}$$

Let d_{NT} be a sequence satisfying

$$d_{NT} = O(\phi_{NT}). \tag{4.6}$$

Theorem 1. *Suppose that Conditions (C1)-(C5) hold, and* $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$ *and* $K_N^{-1}(\log NT)(\log N)^4 = o(1)$. *Suppose also that the algorithm in Section 3.1 converges within a finite number of iterations. Then, for any* t *there is a stochastically bounded sequence* $\delta_{N,jt}$ *such that as* $N \to \infty$,

$$\sqrt{N}(\boldsymbol{\Sigma}_{Nt}^0)^{-1/2}(\widehat{f}_t - f_t^0 - d_{NT}\delta_{N,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_{J+1}),$$

where $\delta_{N,t} = (\delta_{N,jt}, 0 \leq j \leq J)^\intercal$, d_{NT} *is given in (4.6), and* $\mathbf{I}_{J+1}$ *is the* $(J+1) \times (J+1)$ *identity matrix.*

Remark 1: By using the asymptotic normality provided in 1, we can conduct inference for f_t^0 for each t, such as constructing the confidence interval. Note that in the above asymptotic distribution, there is a bias term $d_{NT}\delta_{N,t}$ involved. In order to let the asymptotic bias negligible, we can further assume that $K_N T^{-1} = o(1)$, $K_N^6 N^{-1}(\log NT)^2 = o(1)$, $NK_N^{-2r} = o(1)$ and $r > 3$. By using the cubic splines, which has the order $m = 4$ and letting $r = m = 4$, we need $NK_N^{-8} = o(1)$. If we let $K_N \asymp N^{1/7}$ and $T \asymp N^{\varrho}$ for some constant $\varrho > 1/7$, then the asymptotic bias is negligible and thus we have

$$\sqrt{N}(\boldsymbol{\Sigma}_{Nt}^0)^{-1/2}(\widehat{f}_t - f_t^0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}_{J+1}).$$

Next theorem establishes the rate of convergence of the final estimator $\widehat{g}_j(x_j)$.

Theorem 2. *Suppose that the same conditions as given in Theorem 1 hold. Then, for each* j,

$$\left[\int \{\widehat{g}_j(x_j) - g_j^0(x_j)\}^2 dx_j\right]^{1/2} = O_p(\phi_{NT}) + o_p(N^{-1/2}), \tag{4.7}$$

where ϕ_{NT} *is given in (4.5).*

Remark 2: The orders $\sqrt{K_N/(NT)}$ and K_N^{-r} are from the noise and bias terms for nonparametric estimation, respectively, and the order $K_N^{3/2}N^{-3/4}\sqrt{\log N}$ from the approximation of the Bahadur representation in the quantile regression setting. This says that if the order $K_N = O((NT)^{1/(2r+1)})$ is chosen, and provided $r - \alpha > 1/2$, where $T = O(N^\alpha)$, then the rate in (4.7) is $O_P((NT)^{-r/(2r+1)})$, which is optimal, see for example, Chen and Christensen (2015).

Remark 3. It is possible to develop inferential results for g_j following Chen and Liao (2012) and Chen and Pouzo (2015). As is usual in nonparametric estimation, the weak cross-sectional and temporal dependence does not affect the limiting distribution, and so standard techniques can be applied. In fact, one may conclude the estimation algorithm with a kernel step and demonstrate the oracle efficiency property, Horowitz and Mammen (2011).

5 Covariance estimation and hypothesis testing for the factors

In order to construct the confidence interval we need to estimate Ω_N^0 and Λ_{Nt}^0, since they are unknown. For estimation of Λ_{Nt}^0, if we use its sample analogue, the conditional density $p_i\left(0\,|X_i, f_t\right)$ needs to be estimated. Instead of using this direct way, we use the Powell's kernel estimation idea in Powell (1991), and estimate Λ_{Nt}^0 by

$$\widehat{\Lambda}_{Nt} = (Nh)^{-1}\sum\nolimits_{i=1}^{N} K\left(\frac{y_{it}-\widehat{f}_{ut}-\sum_{j=1}^{J}\widehat{g}_j(X_{ji})\widehat{f}_{jt}}{h}\right)\widehat{G}_i(X_i)\widehat{G}_i(X_i)^{\intercal}, \tag{5.1}$$

where $\widehat{G}_i(X_i)=\{1,\widehat{g}_1(X_{1i}),\ldots,\widehat{g}_J(X_{Ji})\}^{\intercal}$, while $K(\cdot)$ is the uniform kernel $K(u)=2^{-1}I(|u|\leq 1)$ and h is a bandwidth.

First, we show that the estimator $\widehat{\Lambda}_{Nt}$ is a consistent estimator of Λ_{Nt}^0 given in the theorem below.

Theorem 3. *Suppose that the same conditions as given in Theorem 1 hold, and* $h\to 0$, $h^{-1}\phi_{NT}=o(1)$, $h^{-1}N^{-1/2}=O(1)$, *where* ϕ_{NT} *is given in (4.5). Then, we have* $||\widehat{\Lambda}_{Nt}-\Lambda_{Nt}^0||=o_p(1)$.

Moreover, the exact form of Ω_N^0 defined in Condition (C6) is given by

$$\begin{aligned}\Omega_N^0 &= (NT)^{-1}\sum\nolimits_{t=1}^{T} E\left[\left\{\sum\nolimits_{i=1}^{N} G_i^0(X_i)(\tau-I(\varepsilon_{it}<0))\right\}\left\{\sum\nolimits_{i=1}^{N} G_i^0(X_i)(\tau-I(\varepsilon_{it}<0))\right\}^{\intercal}\right]\\ &= \frac{\tau(1-\tau)}{N}\sum\nolimits_{i=1}^{N} E\{G_i^0(X_i)G_i^0(X_i)^{\intercal}\}+(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i\neq j}^{N} E(v_{it}v_{jt}^{\intercal}),\end{aligned}$$

where $v_{it}=G_i^0(X_i)(\tau-I(\varepsilon_{it}<0))$ for $i=1,\ldots,N$. To estimate Ω_N^0, its sample analogue is not consistent. Kernel-based robust estimators that account for heteroskedasticity and cross-sectional correlation (HAC) are developed (Conley, 1999), and are shown to be consistent under a variety of sets of conditions. It requires to use a truncation lag or "bandwidth", which tends to infinity at a slower rate as N. As pointed out by Kiefer and Vogelsang (2005), this is a convenient assumption mathematically to ensure consistency, but it is unrealistic in finite sample studies. Adopting the idea in Kiefer and Vogelsang (2005), we let the bandwidth M be proportional to the sample size n, i.e., $M=bN$ for $b\in(0,1]$, and then we derive the fixed-b asymptotics (Kiefer and Vogelsang; 2005) for the HAC estimator of Ω_N^0 under the quantile setting. The HAC estimator is given as $\widehat{\Omega}_{N,M}=T^{-1}\sum_{t=1}^{T}\widehat{\Omega}_{Nt,M}$, where

$$\widehat{\Omega}_{Nt,M}=\frac{\tau(1-\tau)}{N}\sum\nolimits_{i=1}^{N}\widehat{G}_i(X_i)\widehat{G}_i(X_i)^{\intercal}+N^{-1}\sum\nolimits_{i\neq j}^{N} K^{*}\left(\frac{i-j}{M}\right)\widehat{v}_{it}\widehat{v}_{jt}^{\intercal}, \tag{5.2}$$

where: $\widehat{v}_{it}=\widehat{G}_i(X_i)(\tau-I(\widehat{\varepsilon}_{it}<0))$ for $i=1,\ldots,N$, $\widehat{\varepsilon}_{it}=y_{it}-\widehat{f}_{ut}-\sum_{j=1}^{J}\widehat{g}_j(X_{ji})\widehat{f}_{jt}$, $K^*(u)$ is a symmetric kernel weighting function satisfying $K^*(0)=1$, and $|K^*(u)|\leq 1$, and M

trims the sample autocovariances and acts as a truncation lag. Consistency of $\widehat{\Omega}_{N,M}$ needs that $M \to \infty$ and $M/N \to 0$. The following theorem provides the limiting distribution of $\widehat{\Omega}_{N,M=bN}$ when $M = bN$ for $b \in (0, 1]$.

Next, we will show asymptotic theory for the HAC covariance estimator under a sequence where the smoothing parameter M equals to bN. Let $\Omega^0 = \lim_{N\to\infty} \Omega_N^0$, and Ω^0 can be written as $\Omega^0 = \Upsilon\Upsilon^\intercal$, where Υ is a lower triangular matrix obtained from the Cholesky decomposition of Ω_t^0.

Theorem 4. *Suppose that the same conditions as given in Theorem 1 hold, and $\phi_{NT} N^{1/2} = o(1)$, and $K^{*\prime\prime}(u)$ exists for $u \in [-1, 1]$ and is continuous. Let $M = bN$ for $b \in (0, 1]$. Then as $N \to \infty$,*

$$\widehat{\Omega}_{N,M=bN} \xrightarrow{\mathcal{D}} \Upsilon \int_0^1 \int_0^1 -\frac{1}{b^2} K^{*\prime\prime}\left(\frac{r-s}{b}\right) B_{J+1}(r) B_{J+1}(s)^\intercal dr ds \Upsilon^\intercal,$$

where $B_{J+1}(r) = W_{J+1}(r) - rW_{J+1}(1)$ denotes a $(J + 1) \times 1$ vector of standard Brownian bridges, and $W_{J+1}(r)$ denotes a $(J + 1)$-vector of independent standard Wiener processes where $r \in [0, 1]$.

Theorem 4 establishes the limiting distribution of $\widehat{\Omega}_{N,M=bN}$, although $\widehat{\Omega}_{N,M=bN}$ is an inconsistent estimator of Ω^0. By using the result in Theorem 4, we construct asymptotically pivotal tests involving f_t^0.

Consider testing the null hypothesis H_0: $Rf_t^0 = r$ against the alternative hypothesis H_1: $Rf_t^0 \neq r$, where R is a $q \times (J + 1)$ matrix with rank q and r is a $q \times 1$ vector. We construct an F-type statistic given as

$$F_{Nt,b} = N(R\widehat{f}_t - r)^\intercal \{R\tau(1-\tau)\widehat{\Lambda}_{Nt}^{-1} \widehat{\Omega}_{N,M=bN} \widehat{\Lambda}_{Nt}^{-1} R^\intercal\}^{-1} (R\widehat{f}_t - r)/q.$$

When $q = 1$, we can construct a t-type statistic:

$$T_{Nt,b} = \frac{N^{1/2}(R\widehat{f}_t - r)}{\sqrt{R\tau(1-\tau)\widehat{\Lambda}_{Nt}^{-1} \widehat{\Omega}_{N,M=bN} \widehat{\Lambda}_{Nt}^{-1}\}^{-1} R^\intercal}}.$$

The limiting distributions of $F_{Nt,b}$ and $T_{Nt,b}$ under the null hypothesis are given in the following theorem.

Theorem 5. *Suppose that the same conditions as given in Theorem 1 hold, and $\phi_{NT} N^{1/2} = o(1)$, and $K^{*\prime\prime}(u)$ exists for $u \in [-1, 1]$ and is continuous. Let $M = bN$ for $b \in (0, 1]$. Then under the null hypothesis H_0: $Rf_t^0 = r$, as $N \to \infty$,*

$$F_{Nt,b} \xrightarrow{\mathcal{D}} \{\tau(1-\tau)\}^{-1} W_q(1)^\intercal \left\{\int_0^1 \int_0^1 -\frac{1}{b^2} K^{*\prime\prime}\left(\frac{r-s}{b}\right) B_q(r) B_q(s)^\intercal dr ds\right\}^{-1} W_q(1)/q.$$

If $q = 1$, then as $N \to \infty$,

$$T_{Nt,b} \xrightarrow{\mathcal{D}} \frac{W_1(1)}{\sqrt{\tau(1-\tau)}\sqrt{\int_0^1\int_0^1 -\frac{1}{b^2}K^{*\prime\prime}\left(\frac{r-s}{b}\right)B_1(r)B_1(s)drds}}.$$

Let $\Lambda_t^0 = \lim_{N\to\infty}\Lambda_{Nt}^0$. The limiting distributions of $F_{Nt,b}$ and $T_{Nt,b}$ under the alternative hypothesis H_1: $Rf_t^0 = r + cN^{-1/2}$ are given in the following theorem.

Theorem 6. *Let $\Upsilon_t^* = (R\Lambda_t^{-1}\Omega^0\Lambda_t^{-1}R^\intercal)^{1/2}$. Suppose that the same conditions as given in Theorem 1 hold, and $\phi_{NT}N^{1/2} = o(1)$, and $K^{*\prime\prime}(u)$ exists for $u \in [-1, 1]$ and is continuous. Let $M = bN$ for $b \in (0, 1]$. Then under the alternative hypothesis H_1: $Rf_t^0 = r + cN^{-1/2}$, as $N \to \infty$,*

$$F_{Nt,b} \xrightarrow{\mathcal{D}} \{\tau(1-\tau)\}^{-1}\{\Upsilon_t^{*-1}c + W_q(1)\}^\intercal \times$$
$$\left\{\int_0^1\int_0^1 -\frac{1}{b^2}K^{*\prime\prime}\left(\frac{r-s}{b}\right)B_q(r)B_q(s)^\intercal drds\right\}^{-1}\{\Upsilon_t^{*-1}c + W_q(1)\}/q.$$

If $q = 1$, then as $N \to \infty$,

$$T_{Nt,b} \xrightarrow{\mathcal{D}} \frac{\Upsilon_t^{*-1}c + W_1(1)}{\sqrt{\tau(1-\tau)}\sqrt{\int_0^1\int_0^1 -\frac{1}{b^2}K^{*\prime\prime}\left(\frac{r-s}{b}\right)B_1(r)B_1(s)drds}}.$$

Remark. If $K^*(x)$ is the Bartlett kernel, then

$$\int_0^1\int_0^1 -\frac{1}{b^2}K^{*\prime\prime}\left(\frac{r-s}{b}\right)B_q(r)B_q(s)^\intercal drds$$
$$= \frac{2}{b}\int_0^1 B_q(r)B_q(r)^\intercal dr - \frac{1}{b}\int_0^{1-b}\{B_q(r+b)B_q(r)^\intercal + B_q(r)B_q(r+b)^\intercal\}dr.$$

These results allow one to test whether the factors are zero in a particular time period or not. Our tests are robust to the form of the cross-sectional dependence in the idiosyncratic error.

6 Application

In a series of important papers, Fama and French (hereafter denoted FF), building on earlier work by Banz (1981), Basu (1977), Rosenberg, Reid and Lanstein (1985) and others, demonstrate that there have been large return premia associated with size and value. These size and value return premia are evident in US data for the period covered by the CRSP/Compustat database (FF (1992)), in earlier US data (Davis (1994), and in non-US equity markets (FF (1998), Hodrick, Ng and Sangmueller (1999)). FF (1993,1995,1996,1998) contended that these return premia can be ascribed to a rational asset pricing paradigm in which the size and value characteristics proxy for assets' sensitivities to pervasive sources of risk in the economy. Haugen (1995) and Lakonoshik, Shleifer and Vishny (1994) argued that the observed value and size return premia arise from market inefficiencies rather than from rational risk premia associated with pervasive sources of risk. They argue that these characteristics do not generate enough nondiversifiable risk to justify the observed premia. Similarly, MacKinlay (1995) argues that the return premia are too large relative to the return volatility of the factor portfolios designed to capture these characteristics, and this creates a near-arbitrage opportunity in the FF model. Daniel and Titman (1997) argued that the factor returns associated with the characteristics are partly an artifact of the FF factor model estimation methodology. Hence the accuracy and reliability of FF's estimation procedure is a critical issue in this research controversy. FF (1993) used a simple portfolio sorting approach to estimate their factor model.

In our data analysis, we use all securities from Center for Research in Security Prices (CRSP) which have complete daily return records from 2005 to 2013, and have two-digit Standard Industrial Classification code (from CRSP), market capitalization (from Compustat) and book value (from Compustat) records. We use daily returns in excess of the risk-free return of 337 stocks. We consider the same four characteristic variables as given in Connor, Matthias and Linton (2012), and Fan, Liao and Wang (2016), which are size, value, momentum and volatility. Connor, Matthias and Linton (2012) provided some detailed descriptions of these characteristics. They are calculated using the same method as described in Fan, Liao and Wang (2016).

We fit the quantile regression model (2.1) for each year, so that there are $T = 251$ observations. By taking the same strategy as in Ma and He (2016), we select the number of interior knots L_N by minimizing the Bayesian information criterion (BIC) given as

$$\mathrm{BIC}(L_N) = \log\{(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\rho_\tau(y_{it} - \widehat{f}_{ut} - \sum\nolimits_{j=1}^{J}\widehat{g}_j(X_{ji})\widehat{f}_{jt})\} + \frac{\log(NT)}{2NT}J(L_N + m).$$

For the estimator $\widehat{\Lambda}_{Nt}$ given in (5.1), the optimal order for the bandwidth h is in the order of $N^{-1/5}$. Similar to Ma and He (2016), we let $h = \kappa N^{-1/5}$ in our numerical analysis and take different values for κ. For the estimator $\widehat{\Omega}_{Nt,M=bN}$ given in (5.2), we use different values for b, and use the Bartlett kernel as suggested in Kiefer and Vogelsang (2005).

Figures 1-3 show the plots of the four estimated loading functions for the year of 2009, 2010, 2011, and 2012 at different quantiles $\tau = 0.2$, 0.5 and 0.8. We observe that the estimated loading functions have similar shapes for these four years. Moreover, for the size, value and momentum characteristics, the estimated functions show a clear nonlinear pattern, and at different quantiles, the curves are different for the same characteristic. For example, for the size characteristic, the estimated loading function fluctuates around zero and it has a sharp drop after the value of size variable exceeds certain value at the quantiles $\tau = 0.2$ and 0.8. However, it has a smooth decreasing pattern for the median with $\tau = 0.5$. For the momentum characteristic, the estimated function shows different curves at the three quantiles.

Next, we let $\kappa = 0.5, 1, 1.5$ and $b = 0.2, 0.4, 0.6$, respectively, for calculation of $\widehat{\Lambda}_{Nt}$ and $\widehat{\Omega}_{Nt,M=bN}$. Using the year of 2012, we test for the statistical significance of each factor at each time point, based on the t-type statistic proposed and its distribution given in Theorem 5. Then for each factor, we find the percentage of the t-type statistics that are significant at a 95% confidence level across the 251 time periods. Table 1 shows the annualized standard deviations of the factor returns, the percentage of significant t-type statistics for each factor, and the average p-value at $\tau = 0.5$. We can see that the results for different values of κ and b are consistent. Moreover, all five factors are statistically significant with the average p-value smaller than 0.05.

7 Conclusions and discussion

We have taken for granted that the J factors are present in the sense that

$$\operatorname{p}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} f_{jt}^{0} \neq 0 \tag{7.1}$$

for $j = 1, \ldots, J$. For the factors in our application this is quite a standard assumption, but in some cases one might wish to test this because if this condition fails, then the right hand side of (2.4) is close to zero and this equation can't identify $g_j^0(x_j)$. We outline below a test of the hypothesis (7.1) based on the unstructured additive quantile regression 2.3). A more limited objective is to test whether for a given time period t, $f_{jt} = 0$, which we provide above.

We are interested in testing the hypothesis that

$$H_{0_{A_j}} : \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} h_{jt}(x_j) = 0 \text{ for all } x_j, \tag{7.2}$$

against the general alternative that $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} h_{jt}(x_j) = \mu_j(x_j)$ with $\int \mu_j(x_j)^2 dP_j(x_j) > 0$. We also may be interested in a joint test $H_0 = \cap_{j\in I_J} H_{0_{A_j}}$, where I_J is a set of integers, a subset of $\{1, 2, \ldots, J\}$. These are tests of the presence of a factor.

We let

$$\widehat{\tau}_{j,n,T} = \frac{\int \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{h}_{jt}(x_j)\right)^2 dP_j(x_j) - a_{n,T}}{s_{n,T}}, \tag{7.3}$$

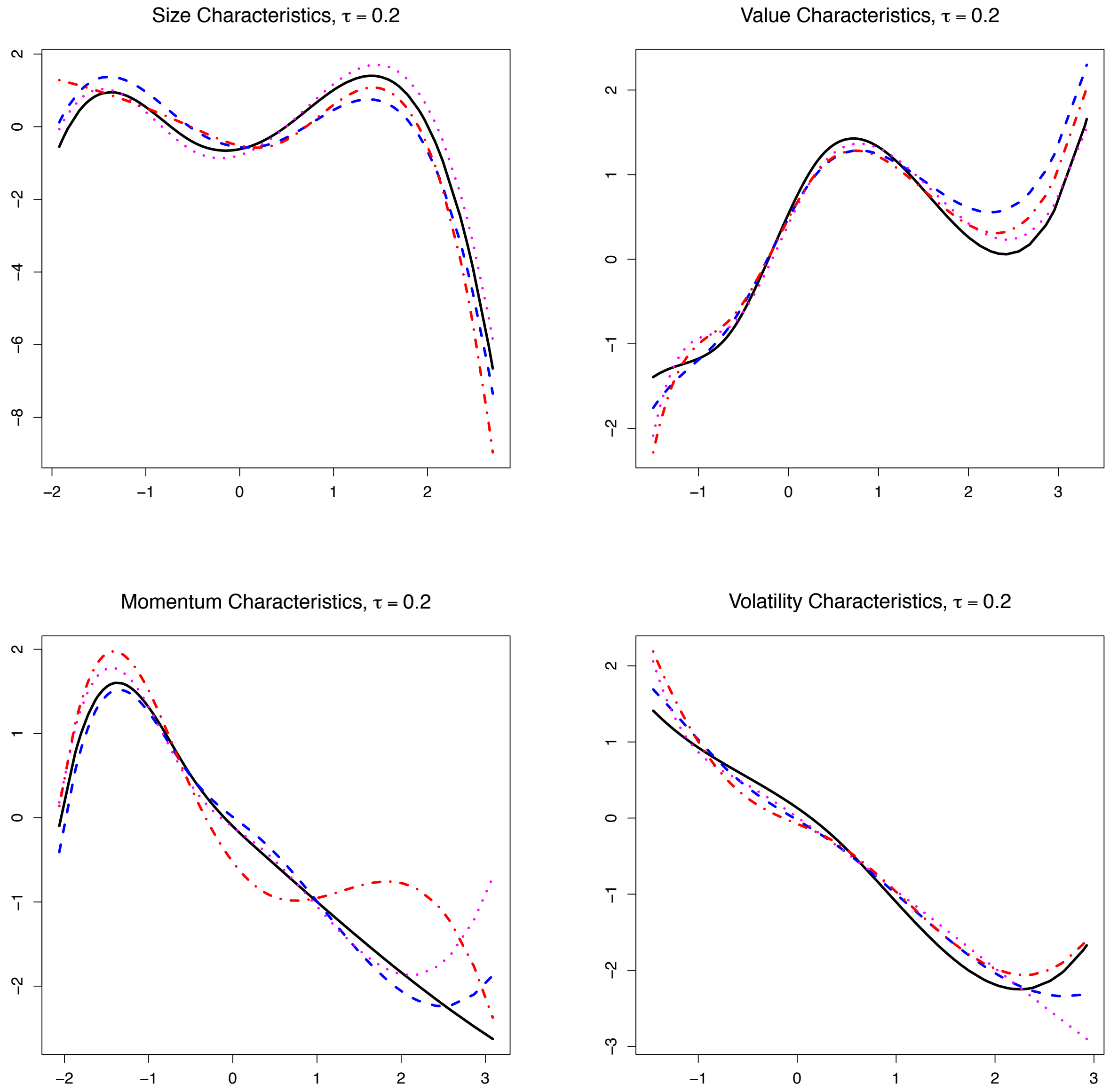
where $\widehat{h}_{jt}(\cdot)$ is the estimated additive component function from the quantile additive model at time t, while $a_{n,T}, s_{n,T}$ are constants to be determined. Under the null hypothesis (7.2) we may show that

$$\widehat{\tau}_{j,n,T} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

Table 1: Factor return statistics at $\tau = 0.5$ for the year of 2012.

(c, b)		Intercept	Size	Value	Momentum	Volatility
	Annualized volatility	0.026	0.026	0.025	0.025	0.026
$(0.5, 0.2)$	% Periods significant	92.00	63.35	65.74	66.14	77.69
	Overall p-value	< 0.001	0.011	0.010	0.011	< 0.001
	Annualized volatility	0.023	0.022	0.022	0.022	0.023
$(0.5, 0.4)$	% Periods significant	93.22	66.93	68.53	69.32	79.28
	Overall p-value	< 0.001	0.006	0.006	0.005	< 0.001
	Annualized volatility	0.020	0.020	0.019	0.019	0.019
$(0.5, 0.6)$	% Periods significant	93.62	72.11	71.71	71.31	81.67
	Overal p-value	< 0.001	0.003	0.003	0.002	< 0.001
	Annualized volatility	0.028	0.032	0.027	0.027	0.029
$(1.0, 0.2)$	% Periods significant	91.63	54.58	61.35	62.55	76.49
	Overall p-value	< 0.001	0.030	0.016	0.017	0.001
	Annualized volatility	0.024	0.027	0.024	0.024	0.025
$(1.0, 0.4)$	% Periods significant	93.23	60.96	65.34	67.73	76.89
	Overall p-value	< 0.001	0.018	0.009	0.008	< 0.001
	Annualized volatility	0.021	0.025	0.021	0.020	0.021
$(1.0, 0.6)$	% Periods significant	93.63	64.94	68.13	70.52	81.27
	Overall p-value	< 0.001	0.010	0.005	0.004	< 0.001
	Annualized volatility	0.030	0.035	0.029	0.029	0.031
$(1.5, 0.2)$	% Periods significant	91.63	51.39	58.17	60.96	75.29
	Overall p-value	< 0.001	0.043	0.020	0.022	0.002
	Annualized volatility	0.026	0.031	0.026	0.025	0.027
$(1.5, 0.4)$	% Periods significant	92.82	56.57	64.94	66.53	75.69
	Overall p-value	< 0.001	0.028	0.013	0.011	< 0.001
	Annualized volatility	0.023	0.027	0.022	0.022	0.022
$(1.5, 0.6)$	% Periods significant	93.63	64.14	66.93	69.32	78.49
	Overall p-value	< 0.001	0.017	0.006	0.005	< 0.001

Figure 1: The plots of the estimated loading functions for the year of 2009 (dotted-dashed red lines), 2010 (dotted magenta lines), 2011 (dashed blue lines), and 2012 (solid black lines) at $\tau = 0.2$.

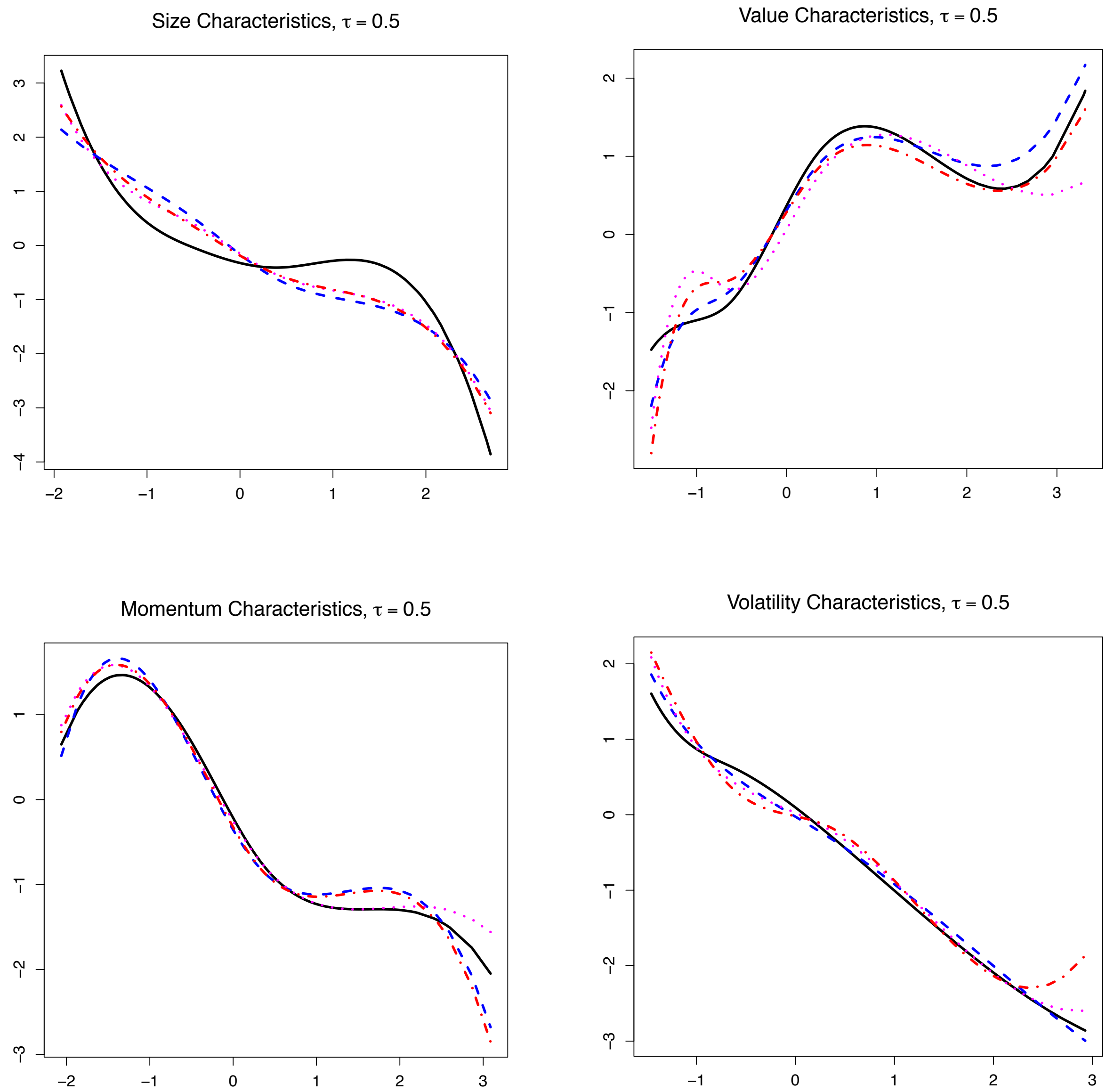


while under the alternative we have $\widehat{\tau}_{j,n,T} \to \infty$ with probability one.

8 Appendix

We first introduce some notations which will be used throughout the Appendix. Let $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote the largest and smallest eigenvalues of a symmetric matrix $\mathbf{A}$, respectively. For an $m \times n$ real matrix A, we denote $\|\mathbf{A}\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |A_{ij}|$. For any vector $\mathbf{a} = (a_1, \ldots, a_n)^\intercal \in \mathbb{R}^n$, denote $||\mathbf{a}||_\infty = \max_{1\le i\le n} |a_i|$. We first study the asymptotic properties of the initial estima-

Figure 2: The plots of the estimated loading functions for the year of 2009 (dotted-dashed red lines), 2010 (dotted magenta lines), 2011 (dashed blue lines), and 2012 (solid black lines) at $\tau = 0.5$.



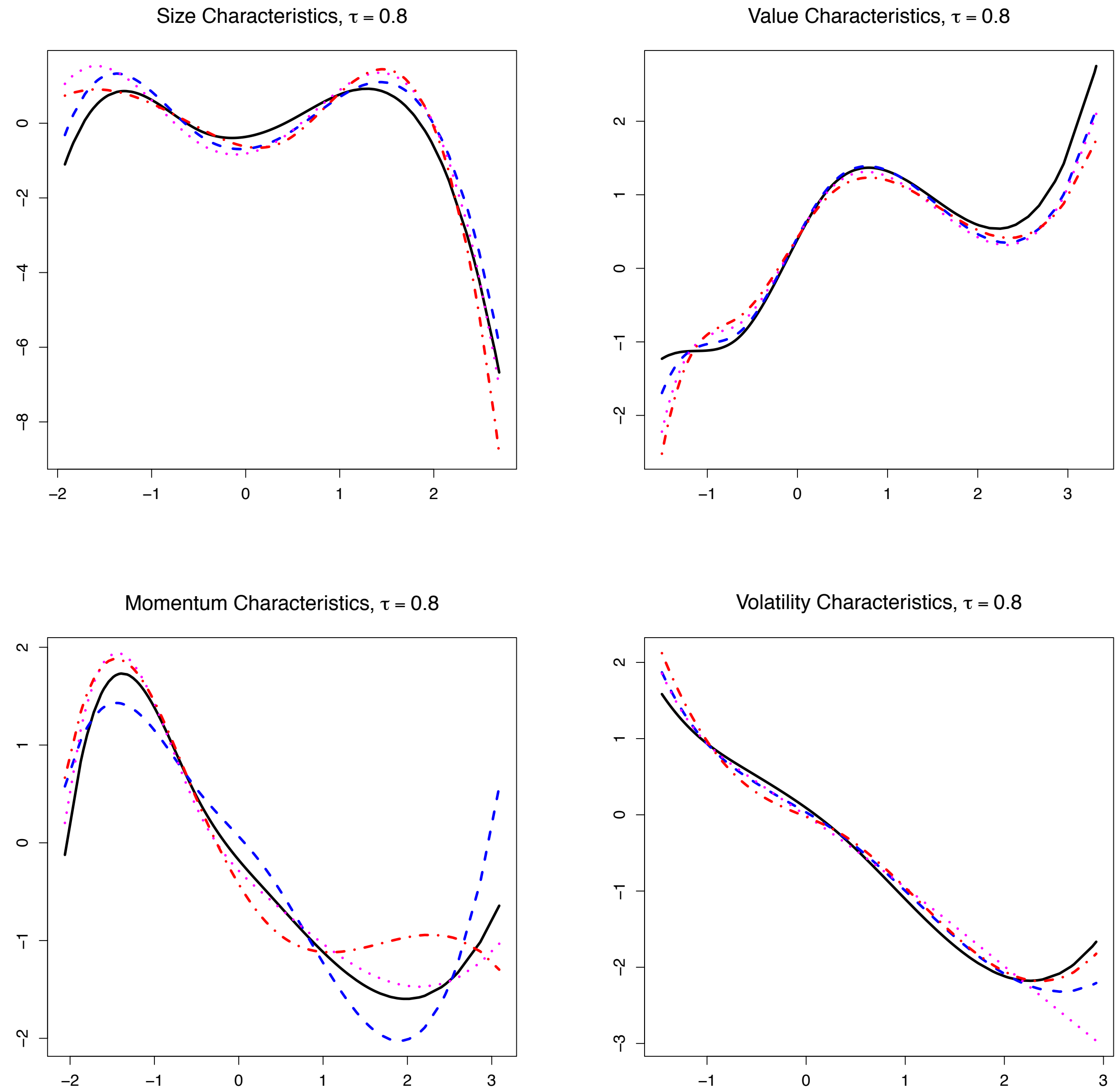
tors $\widehat{g}_j^{[0]}(x_j)$ of $g_j^0(x_j)$. The following theorem gives an asymptotic expression of $\widehat{g}_j^{[0]}(x_j)$ and its convergence rate that will be used in the proofs of Theorems 1 and 2.

Proposition 1. *Let Conditions (C1)-(C4) hold. If, in addition,* $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$ *and* $K_N^{-1}(\log NT)(\log N)^4 = o(1)$, *then for every* $1 \leq j \leq J$,

$$\sup_{x_j \in [a,b]} |\widehat{g}_j^{[0]}(x_j) - g_j^0(x_j)| = O_p(K_N/\sqrt{NT} + K_N^2 N^{-3/4}\sqrt{\log NT} + K_N^{-r}) + o_p(N^{-1/2}),$$

$$\left[\int \{\widehat{g}_j^{[0]}(x_j) - g_j^0(x_j)\}^2 dx_j\right]^{1/2} = O_p(\sqrt{K_N/(NT)} + K_N^{3/2} N^{-3/4}\sqrt{\log NT} + K_N^{-r}) + o_p(N^{-1/2}). \tag{A.1}$$

Figure 3: The plots of the estimated loading functions for the year of 2009 (dotted-dashed red lines), 2010 (dotted magenta lines), 2011 (dashed blue lines), and 2012 (solid black lines) at $\tau = 0.8$.



8.1 Proof of Proposition 1

We first present the following several lemmas which will be used in the proof of Proposition 1. According to the result on page 149 of de Boor (2001), for h^0_{jt} satisfying the smoothness condition given in (C2), there exists $\boldsymbol{\theta}^0_{jt} \in \mathbb{R}^{K_n}$ such that $h^0_{jt}(x_j) = \widetilde{h}^0_{jt}(x_j) + b_{jt}(x_j)$

$$\widetilde{h}^0_{jt}(x_j) = B_j(x_j)^\intercal \boldsymbol{\theta}^0_{jt} \text{ and } \sup_{j,t} \sup_{x_j \in [a,b]} |b_{jt}(x_j)| = O(K_N^{-r}). \tag{A.2}$$

Denote $\widetilde{h}^0_t(x) = \{\widetilde{h}^0_{jt}(x_j), 1 \le j \le J\}^\intercal$, and

$$b_t(x) = \sum\nolimits_{j=1}^{J} h^0_{jt}(x_j) - B(x)^\intercal \boldsymbol{\theta}^0_t$$

where $B(x) = \{B_1(x_1)^\intercal, \ldots, B_J(x_J)^\intercal\}^\intercal$. Then by (A.2), we have

$$\sup_{x\in[a,b]^J} |b_t(x)| = O(K_N^{-r}).$$

Let $\boldsymbol{\theta}_t^0 = (\boldsymbol{\theta}_{1t}^{0\intercal}, \ldots, \boldsymbol{\theta}_{Jt}^{0\intercal})^\intercal$. Then $\mathbb{B}(x)(\widetilde{h}_{ut}, \widetilde{\boldsymbol{\theta}}_t^\intercal)^\intercal = (\widetilde{h}_{ut}, \widetilde{h}_t(x)^\intercal)^\intercal$ and $\mathbb{B}(x)(h_{ut}^0, \boldsymbol{\theta}_t^{0\intercal})^\intercal = (h_{ut}^0, \widetilde{h}_t^0(x)^\intercal)^\intercal$, where $\mathbb{B}(x)$ is defined in (4.2). We introduce some additional notation that were used in Koenker and Bassett (1978), and Horowitz and Lee (2005). Let $d(N) = (1 + JK_N)$. Let $\mathcal{N} = \{1, \ldots, N\}$ and $\mathcal{S}$ denote the collection of all $d(N)$-element subsets of $\mathcal{N}$. Let $\mathbf{M}(s)$ denote the submatrix (subvector) of a matrix (vector) $\mathbf{M}$ with rows (components) indexed by the elements of $s \in \mathcal{S}$. Let $\mathbf{Z} = (Z_1, \ldots, Z_N)^\intercal$, where Z_i is defined in (4.1), and $Y_t = (y_{it}, 1 \le i \le N)^\intercal$. Then $\mathbf{Z}(s)$ is the $d(N) \times d(N)$ matrix, whose rows are Z_i's with $i \in s$, and $Y_t(s)$ is the $d(N) \times 1$ vector, whose elements are y_{it}'s with $i \in s$ for each given t. We first give the Bernstein inequality for a ϕ-mixing sequence, which is used through our proof.

Lemma 1. *Let $\{\xi_i\}$ be a sequence of centered real-valued random variables. Let $S_n = \sum_{i=1}^n \xi_i$. Suppose the sequence has the ϕ-mixing coefficient satisfying $\phi(k) \le \exp(-2ck)$ for some $c > 0$ and $\sup_{i\ge1} |\xi_i| \le M$. Then there is a positive constant C_1 depending only on c such that for all $n \ge 2$*

$$P(|S_n| \ge \varepsilon) \le \exp(-\frac{C_1\varepsilon^2}{v^2 n + M^2 + \varepsilon M(\log n)^2}),$$

where $v^2 = \sup_{i>0}(var(\xi_i) + 2\sum_{j>i} |cov(\xi_i, \xi_j)|)$.

Proof. The result of Lemma 1 is given in Theorem 2 on page 275 of Merlevéde, Peligrad and Rio (2009) when the sequence $\{\xi_i\}$ has the α-mixing coefficient satisfying $\alpha(k) \le \exp(-2ck)$ for some $c > 0$. Thus, this result also holds for the sequence having the ϕ-mixing coefficient satisfying $\phi(k) \le \exp(-2ck)$, since $\alpha(k) \le \phi(k) \le \exp(-2ck)$. □

Lemma 2. *There is a subset $s \in \mathcal{S}$ such that the objective function (3.2) has at least one minimizer of the form $(\widetilde{h}_{ut}, \widetilde{\boldsymbol{\theta}}_t^\intercal)^\intercal = \mathbf{Z}(s)^{-1} Y_t(s)$, and $(\widetilde{h}_{ut}, \widetilde{\boldsymbol{\theta}}_t^\intercal)^\intercal$ is a unique solution to (3.2) almost surely for sufficiently large N.*

Proof. The proof of this lemma is given in Lemma A.2 of Horowitz and Lee (2005). □

We first obtain the Bahadur representation for $\widetilde{\boldsymbol{\vartheta}}_t = (\widetilde{h}_{ut}, \widetilde{\boldsymbol{\theta}}_t^\intercal)^\intercal$ through the following lemmas. To obtain the Bahadur representation for $\widetilde{\boldsymbol{\vartheta}}_t$, we basically extend the result established for the i.i.d. case by Horowitz and Lee (2005) to the mixing distribution by following similar procedures as given in Lemmas A.1-A.7 of Horowitz and Lee (2005), and we also need the results to hold uniformly in t, which requires to apply the Bernstein's inequality for mixing distributions in Lemma 1 and the union bound of probability. Denote $\boldsymbol{\vartheta}_t = (h_{ut}, \boldsymbol{\theta}_t^\intercal)^\intercal$ and $\boldsymbol{\vartheta}_t^0 = (h_{ut}^0, \boldsymbol{\theta}_t^{0\intercal})^\intercal$. Define

$$G_{tN,i}(\boldsymbol{\vartheta}_t) = [\tau - I\{\varepsilon_{it} \le Z_i^\intercal(\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0) - b_t(X_i)\}]Z_i,$$
$$G_{tN,i}^*(\boldsymbol{\vartheta}_t) = [\tau - F_i[\{Z_i^\intercal(\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0) - b_t(X_i)\}|X_i, f_t]]Z_i,$$

where $F_i(\varepsilon|X_i, f_t) = P(\varepsilon_i \le \varepsilon|X_i, f_t)$, and $\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t) = G_{tN,i}(\boldsymbol{\vartheta}_t) - G_{tN,i}^*(\boldsymbol{\vartheta}_t)$.

Lemma 3. *Under Conditions (C1) and (C2), and* $K_N N^{-1}(\log K_N T)(\log N)^4 = o(1)$ *and* $K_N^{-1} = o(1)$, $\sup_{1\le t\le T} ||N^{-1}\sum_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)|| = O_p(K_N^{1/2}N^{-1/2}\sqrt{\log K_N T})$.

Proof. It is easy to see that $E\{N^{-1}\sum_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)\} = \mathbf{0}$. Write $Z_i = (Z_{i,1},\ldots,Z_{i,d(N)})^\intercal$. Let

$$\begin{aligned}\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t) &= [\tau - I\{\varepsilon_{it} \le Z_i^\intercal(\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0) - b_t(X_i)\}]Z_{i,\ell} \\ &\quad - [\tau - F_i[\{Z_i^\intercal(\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0) - b_t(X_i)\}|X_i, f_t]]Z_{i,\ell},\end{aligned}$$

where $\ell = 1,\ldots,d(N)$, so that $\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0) = \{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0), 1 \le \ell \le d(N)\}^\intercal$ and $\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) = [F_i[\{-b_t(X_i)\}|X_i, f_t] - I\{\varepsilon_{it} \le -b_t(X_i)\}]Z_{i,\ell}$. Then for each ℓ,

$$E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t)^2\} = E[Var\{I(\varepsilon_{it} \le -b_t(X_i))|X_i, f_t\}Z_{i,\ell}^2] \asymp E(Z_{i,\ell}^2) \asymp 1,$$

and by Condition (C1), for $i \ne i'$,

$$\begin{aligned}|E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t)\widetilde{G}_{tN,i'\ell}(\boldsymbol{\vartheta}_t)\}| &\le 2\{\phi(|i'-i|)\}^{1/2}[E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t)^2\}E\{\widetilde{G}_{tN,i\ell'}(\boldsymbol{\vartheta}_t)^2\}]^{1/2} \\ &\le c_1 2K_1 e^{-\lambda_1|i'-i|/2},\end{aligned}$$

for some constant $0 < c_1 < \infty$. Hence, by the above results, we have

$$\begin{aligned}&\sup_i [E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t)\}^2 + \sum\nolimits_{i'\ne i}|Cov(\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t), \widetilde{G}_{tN,i'\ell}(\boldsymbol{\vartheta}_t))|] \\ &\le c_2 + \sup_i \sum\nolimits_{i'\ne i} c_1 2K_1 e^{-\lambda_1|i'-i|/2} \le c_2 + c_1 2K_1(1 - e^{-\lambda_1/2})^{-1} \le c_3\end{aligned}$$

for some constants $0 < c_2, c_3 < \infty$. Moreover, $\sup_i|\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t)| \le c_4 K_N^{1/2}$ for some constant $0 < c_4 < \infty$. Thus, by the Bernstein's inequality in Lemma 1, we have for N sufficiently large and $K_N N^{-1}(\log K_N T)(\log N)^4 = o(1)$,

$$\begin{aligned}&P\left(|N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t)| \ge aN^{-1/2}\sqrt{\log K_N T}\right) \\ &\le \exp\left(-\frac{C_1 a^2 N(\log K_N T)}{c_3 N + c_4^2 K_N + aN^{1/2}\sqrt{\log K_N T}\, c_4 K_N^{1/2}(\log N)^2}\right) \le (K_N T)^{-C_1 a^2/(3c_3)}.\end{aligned}$$

Then by the union bound of probability, we have

$$P\left(\sup_t \sup_\ell |N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t)| \ge aN^{-1/2}\sqrt{\log K_N T}\right) \le d(N)T(K_N T)^{-C_1 a^2/(3c_3)}.$$

Therefore,

$$\begin{aligned}P\left(\sup_{1\le t\le T} ||N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)|| \ge aK_N^{1/2}N^{-1/2}\sqrt{\log K_N T}\right) &\le d(N)T(K_N T)^{-C_1 a^2/(3c_3)} \\ &\le (1 + JK_N)T(K_N T)^{-2}.\end{aligned}$$

by taking a large enough. The proof is complete. □

Lemma 4. $\sup_{1\le t\le T} ||N^{-1}\sum_{i=1}^N G_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t)|| = O_{a.s.}(K_N^{3/2}N^{-1})$.

Proof. The proof of this lemma follows the same procedure as in Lemma A.4 of Horowitz and Lee (2005) by using the result in (A.9) which holds uniformly in $t = 1, ..., T$. □

Lemma 5. *Under Conditions (C1) and (C2), and* $K_N^2 N^{-1}(\log NT)^2(\log N)^8 = o(1)$ *and* $K_N^{-1} = o(1)$,

$$\sup_{1\le t\le T} \sup_{||\boldsymbol{\vartheta}_t-\boldsymbol{\vartheta}_t^0||\le CK_N^{1/2}N^{-1/2}} ||N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t) - N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)|| \\ = O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}).$$

Proof. Let $B_N = \{\boldsymbol{\vartheta}_t : ||\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0|| \le CK_N^{1/2}N^{-1/2}\}$. By taking the same strategy as given in Lemma A.5 of Horowitz and Lee (2005), we cover the ball B_N with cubes $\mathcal{C} = \{\mathcal{C}(\boldsymbol{\vartheta}_{t,v})\}$, where $\mathcal{C}(\boldsymbol{\vartheta}_{t,v})$ is a cube containing $(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0)$ with sides of $C\{d(N)/N^5\}^{1/2}$ such that $\boldsymbol{\vartheta}_{t,v} \in B_N$. Then the number of the cubes covering the ball B_N is $V = (2N^2)^{d(N)}$. Moreover, we have $||(\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0) - (\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0)|| \le C\{d(N)/N^{5/2}\}$ for any $\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0 \in \mathcal{C}(\boldsymbol{\vartheta}_{t,v})$, where $v = 1, \ldots, V$. First we can decompose

$$\sup_{\boldsymbol{\vartheta}_t\in B_N} ||N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t) - N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)|| \\ \le \max_{1\le v\le V} \sup_{(\boldsymbol{\vartheta}_t-\boldsymbol{\vartheta}_t^0)\in\mathcal{C}(\boldsymbol{\vartheta}_{t,v})} ||N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t) - N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_{t,v})|| \\ + \max_{1\le v\le V} ||N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_{t,v}) - N^{-1}\sum\nolimits_{i=1}^N \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)|| \\ = \Delta_{tN,1} + \Delta_{tN,2} \tag{A.3}$$

Let $\gamma_N = C\{d(N)/n^{5/2}\}$. By the same argument as given in the proof of Lemma A.5 in Horowitz and Lee (2005), we have

$$\Delta_{tN,1} \le \max_{1\le v\le V} |\Gamma_{tN,1v}| + \max_{1\le v\le V} |\Gamma_{tN,2v}|, \tag{A.4}$$

where

$$\Gamma_{tN,1v} = N^{-1}\sum\nolimits_{i=1}^N ||Z_i|| \left[F_i[Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i) + ||Z_i||\gamma_N|X_i, f_t]\right. \\ \left. -F_i[Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i) - ||Z_i||\gamma_N|X_i, f_t]\right],$$

$$\Gamma_{tN,2v} = N^{-1}\sum\nolimits_{i=1}^N \Gamma_{tN,2v,i} = N^{-1}\sum\nolimits_{i=1}^N ||Z_i|| \left[[I\{\varepsilon_{it} \le Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i) + ||Z_i||\gamma_N\}\right. \\ - F_i\{Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i) + ||Z_i||\gamma_N|X_i, f_t\}] \\ \left. -[I\{\varepsilon_{it} \le Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i)\} - F_i\{Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i)|X_i, f_t\}]\right].$$

By Condition (C2), we have for some constants $0 < c_1, c_2 < \infty$,

$$\sup_{1\le t\le T} \max_{1\le v\le V} |\Gamma_{tN,1v}| \le c_1\gamma_N \max_{1\le i\le N} ||Z_i||||Z_i|| \le c_2\{d(N)/N^{5/2}\}K_N = O(K_N^2 N^{-5/2}). \tag{A.5}$$

Next we will show the convergence rate for $\max_{1\le v\le V} |\Gamma_{tN,2v}|$. It is easy to see that $E(\Gamma_{tN,2v,i}) = 0$. Also $|\Gamma_{tN,2v,i}| \le 4||Z_i|| \le c_1 K_N^{1/2}$ for some constant $0 < c_1 < \infty$. Moreover,

$$E\left[||Z_i||I\{\varepsilon_{it} \le Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i) + ||Z_i||\gamma_N\} - I\{\varepsilon_{it} \le Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i)\}\right]^2 \\ \asymp E\{||Z_i||^2||Z_i||\gamma_N\} \le c_2^*\gamma_N K_N^{1/2} \le c_2 K_N^{3/2} N^{-5/2},$$

for some constants $0 < c_2^* < c_2 < \infty$. Hence $E(\Gamma_{tN,2v,i})^2 \le c_2 K_N^{3/2} N^{-5/2}$. By Condition (C1), we have for $i \ne j$,

$$|E(\Gamma_{tN,2v,i}\Gamma_{tN,2v,j})| \le 2\phi(|j-i|)^{1/2}\{E(\Gamma^2_{tN,2v,i})E(\Gamma^2_{tN,2v,j})\}^{1/2} \le 2c_2\phi(|j-i|)K_N^{3/2}N^{-5/2}.$$

Hence

$$\begin{aligned}
&E(\Gamma_{tN,2v,i})^2 + 2\sum\nolimits_{j>i}|E(\Gamma_{tN,2v,i}\Gamma_{tN,2v,j})| \\
&\le c_2 K_N^{3/2}N^{-5/2} + 4c_2\sum\nolimits_{k=1}^{N} K_1 e^{-\lambda_1 k/2}K_N^{3/2}N^{-5/2} \\
&\le c_2 K_N^{3/2}N^{-5/2}(1+4K_1(1-e^{-\lambda_1/2})^{-1}) = c_3 K_N^{3/2}N^{-5/2},
\end{aligned}$$

where $c_3 = c_2(1+4K_1(1-e^{-\lambda_1/2})^{-1})$. By Condition (C1), for each fixed t, the sequence $\{(X_i, f_t, \varepsilon_{it}), 1 \le i \le N\}$ has the ϕ-mixing coefficient $\phi(k) \le K_1 e^{-\lambda_1 k}$ for $K_1, \lambda_1 > 0$. Thus, by the Bernstein's inequality given in Lemma 1, we have for N sufficiently large,

$$\begin{aligned}
&P\left(|\Gamma_{tN,2v}| \ge aK_N^{3/2}N^{-1}(\log NT)^3\right) \\
&\le \exp\Big(-\frac{C_1(aK_N^{3/2}(\log NT)^3)^2}{c_3K_N^{3/2}N^{-5/2}N + c_1^2K_N + aK_N^{3/2}(\log NT)^3c_1K_N^{1/2}\log(N)^2}\Big) \le (NT)^{-c_4a^2K_N}
\end{aligned}$$

for some constant $0 < c_4 < \infty$. By the union bound of probability, we have

$$\begin{aligned}
&P\left(\sup_{1\le t\le T}\max_{1\le v\le V}|\Gamma_{tN,2v}| \ge aK_N^{3/2}N^{-1}(\log NT)^3\right) \\
&\le (2N^2)^{d(N)}T(NT)^{-c_4a^2K_N} \le 2^{d(N)}N^{2(1+JK_N)-c_4a^2K_N}T^{1-c_4a^2K_N}.
\end{aligned}$$

Hence, taking a large enough, one has

$$P\left(\sup_{1\le t\le T}\max_{1\le v\le V}|\Gamma_{tN,2v}| \ge aK_N^{3/2}N^{-1}(\log N)^3\right) \le 2^{K_N}N^{-K_N}T^{-K_N}.$$

Then we have

$$\sup_{1\le t\le T}\max_{1\le v\le V}|\Gamma_{tN,2v}| = O_p\{K_N^{3/2}N^{-1}(\log NT)^3\}. \tag{A.6}$$

Next we will show the convergence rate for $\Delta_{tN,2}$. Let $\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})$ be the ℓ^{th} element in $\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_{t,v}) - \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)$ for $\ell = 1,\ldots,d(N)$. It is easy to see that $E\{\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})\} = 0$. Also $|\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})| \le 4|Z_{i\ell}| \le c_1K_N^{1/2}$ for some constant $0 < c_1 < \infty$. Moreover,

$$\begin{aligned}
&E\left[[I\{\varepsilon_{it} \le Z_i^\intercal(\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0) - b_t(X_i)\} - I\{\varepsilon_{it} \le -b_t(X_i)\}]Z_{i\ell}\right]^2 \\
&\le c_1'||\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0||K_N^{1/2} \le c_1'CK_N^{1/2}N^{-1/2}K_N^{1/2} = c_1'CK_NN^{-1/2}
\end{aligned}$$

for some constant $0 < c_1' < \infty$. Hence $E(\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v}))^2 \le c_1'CK_NN^{-1/2}$. By Condition (C1), we have for $i \ne j$,

$$|E(\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})\widetilde{g}_{tN,j,\ell}(\boldsymbol{\vartheta}_{t,v})| \le 4\phi(|j-i|)^{1/2}\{E(\Gamma^2_{tN,2v,i})E(\Gamma^2_{tN,2v,j})\}^{1/2}.$$

Hence

$$
\begin{aligned}
&E(\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v}))^2 + 2\sum\nolimits_{j>i} |E(\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})\widetilde{g}_{tN,j,\ell}(\boldsymbol{\vartheta}_{t,v})| \\
&\le c_1' C K_N N^{-1/2} + 4\sum\nolimits_{k=1}^{N} K_1 e^{-\lambda_1 k/2} c_1' C K_N N^{-1/2} \\
&\le c_1' C K_N N^{-1/2}(1 + 4K_1(1-e^{-\lambda_1/2})^{-1}) = c_2 K_N N^{-1/2},
\end{aligned}
$$

where $c_2 = c_1' C(1+4K_1(1-e^{-\lambda_1/2})^{-1})$. Thus, by the Bernstein's inequality given in Lemma 1 and $K_N^2 N^{-1}(\log NT)^2(\log N)^8 = o(1)$, we have for N sufficiently large,

$$
\begin{aligned}
&P\left(|N^{-1}\sum\nolimits_{i=1}^{N} \widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})| \ge aK_N N^{-3/4}\sqrt{\log NT}\right) \\
&\le \exp\Big(-\frac{C_1(aK_N N^{1/4}\sqrt{\log NT})^2}{c_2 K_N N^{-1/2}N + c_1^2 K_N + aK_N N^{1/4}(\log NT)^{1/2} c_1 K_N^{1/2}(\log N)^2}\Big) \le (NT)^{-c_3 a^2 K_N} \quad \text{(A.7)}
\end{aligned}
$$

for some constant $0 < c_3 < \infty$. By the union bound of probability, we have

$$
P\left(\sup_{1\le t\le T}\sup_{1\le \ell\le d(N)} |N^{-1}\sum\nolimits_{i=1}^{N} \widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})| \ge aK_N N^{-3/4}\sqrt{\log NT}\right) \le d(N)T(NT)^{-c_3 a^2 K_N}.
$$

Hence,

$$
\begin{aligned}
&P\left(\sup_{1\le t\le T} ||N^{-1}\sum\nolimits_{i=1}^{N} \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_{t,v}) - N^{-1}\sum\nolimits_{i=1}^{N} \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)|| \ge aK_N^{3/2} N^{-3/4}\sqrt{\log NT}\right) \\
&\le d(N)T(NT)^{-c_3 a^2 K_N}.
\end{aligned}
$$

By the union bound of probability again, we have

$$
P\left(\sup_{1\le t\le T} |\Delta_{tN,2}| \ge aK_N^{3/2} N^{-3/4}\sqrt{\log NT}\right) \le (2N^2)^{d(N)} d(N)T(NT)^{-c_3 a^2 K_N}.
$$

Hence, taking a large enough, one has

$$
P\left(\sup_{1\le t\le T} |\Delta_{tN,2}| \ge aK_N^{3/2} N^{-3/4}\sqrt{\log NT}\right) \le 2^{K_N} K_N N^{-K_N} T^{-K_N}.
$$

Then we have

$$
\sup_{1\le t\le T} |\Delta_{tN,2}| = O_p\{K_N^{3/2} N^{-3/4}\sqrt{\log NT}\}. \quad \text{(A.8)}
$$

Therefore, by (A.3), (A.4), (A.5), (A.6) and (A.8), we have

$$
\begin{aligned}
&\sup_{1\le t\le T}\sup_{\boldsymbol{\vartheta}_t\in B_N} ||N^{-1}\sum\nolimits_{i=1}^{N} \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t) - N^{-1}\sum\nolimits_{i=1}^{N} \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)|| \\
&= O_p\{K_N^2 N^{-5/2} + K_N^{3/2} N^{-1}(\log NT)^3 + K_N^{3/2} N^{-3/4}\sqrt{\log NT}\} \\
&= O_p(K_N^{3/2} N^{-3/4}\sqrt{\log NT}).
\end{aligned}
$$

□

Let $\Psi_{Nt} = N^{-1}\sum_{i=1}^{N} p_i\left(0\,|X_i, f_t\right) Z_i Z_i^{\intercal}$. By the same reasoning as the proofs for (ii) of Lemma A.7 in Ma and Yang (2011), we have with probability approaching 1, as $N \to \infty$, there exist constants $0 < C_1 \le C_2 < \infty$ such that

$$C_1 \le \lambda_{\min}(\Psi_{Nt}) \le \lambda_{\max}(\Psi_{Nt}) \le C_2, \tag{A.9}$$

uniformly in $t = 1, ..., T$.

Lemma 6. *Under Conditions (C2) and (C3), as* $N \to \infty$*,*

$$\Psi_{Nt}^{-1} G_{tN,i}^{*}(\boldsymbol{\vartheta}_t) = -(\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0) + N^{-1}\Psi_{Nt}^{-1}\sum_{i=1}^{N} p_i\left(0\,|X_i, f_t\right) Z_i b_t(X_i) + R_{Nt}^{*},$$

where $||R_{Nt}^{*}|| \le C^{*}\{K_N^{1/2}||\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0||^2 + K_N^{-1/2-2r}\}$ *for some constant* $0 < C^{*} < \infty$*, uniformly in* t*.*

Lemma 7. *Under Condition (C2),*

$$\sup_{1\le t\le T} ||N^{-1}\Psi_{Nt}^{-1}\sum_{i=1}^{N} p_i\left(0\,|X_i, f_t\right) Z_i b_t(X_i)|| = O(K_N^{-r}).$$

Proof. The proofs of Lemmas 6 and 7 follow the same procedure as in Lemmas A.6-A.7 of Horowitz and Lee (2005) by using the results (A.2) and (A.9) □

Lemma 8. *Under Conditions (C1)-(C3), and* $K_N^3 N^{-1} = o(1)$*,* $K_N^2 N^{-1}(\log NT)^2(\log N)^8 = o(1)$ *and* $K_N^{-r+1}(\log T) = o(1)$*,*

$$\widetilde{\boldsymbol{\vartheta}}_t - \boldsymbol{\vartheta}_t^0 = D_{Nt,1} + D_{Nt,2} + R_{Nt}, \tag{A.10}$$

where

$$D_{Nt,1} = \left[N^{-1}\sum_{i=1}^{N} p_i\left(0\,|X_i, f_t\right) Z_i Z_i^{\intercal}\right]^{-1}\left[N^{-1}\sum_{i=1}^{N} Z_i(\tau - I(\varepsilon_{it} < 0))\right], \tag{A.11}$$

$$D_{Nt,2} = \Psi_{Nt}^{-1}\left[N^{-1}\sum_{i=1}^{N} Z_i\{p_i\left(0\,|X_i, f_t\right)\sum_{j=1}^{J} b_{jt}(X_{ji})\}\right], \tag{A.12}$$

uniformly in t*, and the remaining term* R_{Nt} *satisfies*

$$\begin{aligned}\sup_{1\le t\le T} ||R_{Nt}|| &= O_p(K_N^{3/2}N^{-1} + K_N^{3/2}N^{-3/4}\sqrt{\log NT} + K_N^{1/2-2r} + N^{-1/2}K_N^{-r/2+1/2}\sqrt{\log K_N T}) \\ &= O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT} + K_N^{1/2-2r}) + o_p(N^{-1/2}).\end{aligned}$$

Proof. By Lemma 6, we have

$$\widetilde{\boldsymbol{\vartheta}}_t - \boldsymbol{\vartheta}_t^0 = N^{-1}\Psi_{Nt}^{-1}\sum_{i=1}^{N} p_i\left(0\,|X_i, f_t\right) Z_i b_t(X_i) - \Psi_{Nt}^{-1} G_{tN,i}^{*}(\widetilde{\boldsymbol{\vartheta}}_t) + R_{Nt}^{*}.$$

Moreover,

$$\Psi_{Nt}^{-1} G_{tN,i}^{*}(\widetilde{\boldsymbol{\vartheta}}_t) = \Psi_{Nt}^{-1} G_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t) - \Psi_{Nt}^{-1}\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0) - \Psi_{Nt}^{-1}[\widetilde{G}_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t) - \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)].$$

Thus,

$$\widetilde{\boldsymbol{\vartheta}}_t - \boldsymbol{\vartheta}_t^0 = \Psi_{Nt}^{-1}N^{-1}\sum_{i=1}^{N}\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0) + \Psi_{Nt}^{-1}N^{-1}\sum_{i=1}^{N} p_i\left(0\,|X_i, f_t\right) Z_i b_t(X_i) + R_{Nt}^{**}, \tag{A.13}$$

where

$$R_{Nt}^{**} = -\Psi_{Nt}^{-1} N^{-1} \sum\nolimits_{i=1}^{N} G_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t) + \Psi_{Nt}^{-1} N^{-1} \sum\nolimits_{i=1}^{N} [\widetilde{G}_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t) - \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)] + R_{Nt}^{*}. \qquad \text{(A.14)}$$

By Lemmas 4 and 5 and (A.9), we have

$$\begin{aligned}
\sup_{1\le t\le T} ||R_{Nt}^{**}|| &\le \sup_{1\le t\le T} ||\Psi_{Nt}^{-1}|| \sup_{1\le t\le T} ||N^{-1} \sum\nolimits_{i=1}^{N} G_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t)|| \\
&\quad + \sup_{1\le t\le T} ||\Psi_{Nt}^{-1}|| \sup_{1\le t\le T} ||N^{-1} \sum\nolimits_{i=1}^{N} [\widetilde{G}_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t) - \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)]|| + \sup_{1\le t\le T} ||R_{Nt}^{*}|| \\
&= O_p(K_N^{3/2} N^{-1} + (K_N^2 N)^{-3/4} \sqrt{\log NT} + K_N^{1/2-2r}).
\end{aligned}$$

Define $\overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) = \{\tau - I(\varepsilon_{it} \le 0)\} Z_{i,\ell}$ and $\overline{G}_{tN,i}(\boldsymbol{\vartheta}_t^0) = \{\overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0), 1 \le \ell \le d(N)\}$. Then $E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)\} = 0$. Moreover,

$$E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)\}^2 \le E\left[I\{\varepsilon_{it} \le -b_t(X_i)\} - I\{\varepsilon_{it} \le 0\} Z_{i,\ell}\right]^2 \le C K_N^{-r}$$

for some constant $0 < C < \infty$, and by Condition (C1), we have

$$\begin{aligned}
&E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)\}\{\widetilde{G}_{tN,i'\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i'\ell}(\boldsymbol{\vartheta}_t^0)\} \\
&\le 2 \times 4^2 \{\phi(|i'-i|)\}^{1/2} [E\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)\}^2 E\{\widetilde{G}_{tN,i'\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i'\ell}(\boldsymbol{\vartheta}_t^0)\}^2]^{1/2} \\
&\le C' K_1 e^{-\lambda_1 |i'-i|/2} K_N^{-r}.
\end{aligned}$$

Hence, by the above results, we have

$$\begin{aligned}
&E[N^{-1} \sum\nolimits_{i=1}^{N} \{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)\}]^2 \\
&\le N^{-1} C K_N^{-r} + N^{-2} \sum\nolimits_{i \ne i'} C' K_1 e^{-\lambda_1 |i'-i|} K_N^{-r} \\
&\le C N^{-1} K_N^{-r} + C' K_1 N^{-2} N (1 - e^{-\lambda_1/2})^{-1} K_N^{-r} \le C'' N^{-1} K_N^{-r},
\end{aligned}$$

for some constant $0 < C'' < \infty$. Thus

$$\begin{aligned}
E||N^{-1} \sum\nolimits_{i=1}^{N} \{\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)\}||^2 &= \sum\nolimits_{\ell=1}^{d(N)} E[N^{-1} \sum\nolimits_{i=1}^{N} \{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)\}]^2 \\
&\le C''(1 + J K_N) N^{-1} K_N^{-r}.
\end{aligned}$$

Therefore, by the Bernstein's inequality and the union bound of probability, we have

$$\sup_{1\le t\le T} ||N^{-1} \sum\nolimits_{i=1}^{N} \{\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0) - \overline{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)\}|| = O_p(N^{-1/2} K_N^{-r/2+1/2} \sqrt{\log K_N T}). \qquad \text{(A.15)}$$

Therefore, by (A.13), (A.14) and (A.15), we have $\widetilde{\boldsymbol{\vartheta}}_t - \boldsymbol{\vartheta}_t^0 = D_{Nt,1} + D_{Nt,2} + R_{Nt}$, where

$$\sup_{1\le t\le T} ||R_{Nt}|| = O_p(K_N^{3/2} N^{-1} + (K_N^2 N)^{-3/4} \sqrt{\log NT} + K_N^{1/2-2r} + N^{-1/2} K_N^{-r/2+1/2} \sqrt{\log K_N T}).$$

□

Proof of Proposition 1. By (A.10) in Lemma 8, we have

$$\widetilde{h}_{jt}(x_j) - \widetilde{h}^0_{jt}(x_j) = 1^{\intercal}_{j+1}\mathbb{B}(x)(D_{Nt,1} + D_{Nt,2}) + 1^{\intercal}_{j+1}\mathbb{B}(x)R_{Nt},$$

and

$$\sup_{1\le t\le T}\{N^{-1}\sum\nolimits_{i=1}^{N}(1^{\intercal}_{j+1}\mathbb{B}(X_i)R_{Nt})^2\}^{1/2} \le \sup_{1\le t\le T}||R_{Nt}||[\lambda_{\max}\{N^{-1}\sum\nolimits_{i=1}^{N}B_j(X_{ji})B_j(X_{ji})^{\intercal}\}]^{1/2}$$
$$= O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT} + K_N^{1/2-2r}) + o_p(N^{-1/2}),$$

$$\begin{aligned}&\sup_{1\le t\le T}\sup_{x\in[a,b]^J}|1^{\intercal}_{j+1}\mathbb{B}(x)R_{Nt}|\\ &\le \sup_{x\in[a,b]^J}||\mathbb{B}(x)^{\intercal}1_{j+1}||\sup_{1\le t\le T}||R_{Nt}||\\ &= O(K_N^{1/2})O_p(K_N^{3/2}N^{-1} + K_N^{3/2}N^{-3/4}\sqrt{\log NT} + K_N^{1/2-2r} + N^{-1/2}K_N^{-r/2+1/2}\sqrt{\log K_NT})\\ &= O_p(K_N^2N^{-3/4}\sqrt{\log NT} + K_N^{1-2r}) + o_p(N^{-1/2}),\end{aligned}$$

by the assumption that $K_N^4N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$ and $r > 2$. Since $h^0_{jt}(x_j) = \widetilde{h}^0_{jt}(x_j) + b_{jt}(x_j)$, then we have

$$\widetilde{h}_{jt}(x_j) - h^0_{jt}(x_j) = 1^{\intercal}_{j+1}\mathbb{B}(x)(D_{Nt,1} + D_{Nt,2}) - b_{jt}(x_j) + 1^{\intercal}_{j+1}\mathbb{B}(x)R_{Nt}.$$

Also by (A.2), we have $\sup_{1\le t\le T}\sup_{x\in[a,b]^J}\left|1^{\intercal}_{j+1}\mathbb{B}(x)D_{Nt,2}\right| = O_p(K_N^{-r})$. Then $\widetilde{h}_{jt}(x_j) - h^0_{jt}(x_j)$ can be written as

$$\widetilde{h}_{jt}(x_j) - h^0_{jt}(x_j) = 1^{\intercal}_{j+1}\mathbb{B}(x)D_{Nt,1} + \eta_{N,jt}(x_j), \tag{A.16}$$

where the remaining term $\eta_{N,jt}(x_j)$ satisfies

$$\sup_{1\le t\le T}[N^{-1}\sum\nolimits_{i=1}^{N}\{\eta_{N,jt}(X_{ji})\}^2]^{1/2} = O_p(K_N^{-r}) + O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}), \tag{A.17}$$

$$\begin{aligned}&\sup_{1\le t\le T}\{\int\eta_{N,jt}(x_j)^2dx_j\}^{1/2} = O_p(K_N^{-r}) + O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}),\\ &\sup_{1\le t\le T}\sup\nolimits_{x_j\in[a,b]}|\eta_{N,jt}(x_j)| = O_p(K_N^{-r}) + O_p(K_N^2N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}).\end{aligned} \tag{A.18}$$

Moreover, by Berntein's inequality, we have $\sup_{1\le t\le T}||D_{Nt,1}|| = O_p(\sqrt{K_N/N}\sqrt{\log K_NT})$. Hence,

$$\begin{aligned}\sup_{1\le t\le T}\sup_{x\in[a,b]^J}|1^{\intercal}_{j+1}\mathbb{B}(x)D_{Nt,1}| &= O_p(\sqrt{\log K_NT}K_N/\sqrt{N}),\\ \sup_{1\le t\le T}\{N^{-1}\sum\nolimits_{i=1}^{N}(1^{\intercal}_{j+1}\mathbb{B}(X_i)D_{Nt,1})^2\}^{1/2} &= O_p(\sqrt{\log K_NT}\sqrt{K_N/N}).\end{aligned} \tag{A.19}$$

Therefore, by (A.16), (A.17), (A.18) and (A.19), we have

$$\begin{aligned}&\sup_{1\le t\le T}N^{-1}\sum\nolimits_{i=1}^{N}\{\widetilde{h}_{jt}(X_{ji}) - h^0_{jt}(X_{ji})\}^2 = O_p((\log K_NT)K_N/N + N^{-2r}),\\ &\sup_{1\le t\le T}\sup\nolimits_{x_j\in[a,b]}|\widetilde{h}_{jt}(x_j) - h^0_{jt}(x_j)| = O_p(\sqrt{\log K_NT}K_NN^{-1/2} + K_N^{-r}).\end{aligned} \tag{A.20}$$

Moreover, by $c_h \le N^{-1}\sum_{i=1}^N h_{jt}^0(X_{ji})^2 \le C_h$ almost surely given in Condition (C3) and the above result, we have with probability approaching 1, as $N\to\infty$, $c_h \le N^{-1}\sum_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2 \le C_h$. By (A.16), we have with probability approaching 1, as $N\to\infty$,

$$\begin{aligned}
&1/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2} - 1/\sqrt{N^{-1}\sum\nolimits_{i=1}^N h_{jt}^0(X_{ji})^2}\\
&= C'\{N^{-1}\sum\nolimits_{i=1}^N h_{jt}^0(X_{ji})^2 - N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2\}\\
&= C'N^{-1}\sum\nolimits_{i=1}^N \{\widetilde{h}_{jt}(X_{ji}) - h_{jt}^0(X_{ji})\}h_{jt}^0(X_{ji})\\
&= C'N^{-1}\sum\nolimits_{i=1}^N 1_{j+1}^\intercal \mathbb{B}(x)D_{Nt,1}h_{jt}^0(X_{ji}) + \varrho_{tN} \qquad \text{(A.21)}
\end{aligned}$$

for some constant $0<C'<\infty$, where $\varrho_{tN} = C'N^{-1}\sum_{i=1}^N \eta_{N,jt}(X_{ji})h_{jt}^0(X_{ji})$. Moreover by (A.17),

$$\begin{aligned}
\sup_{1\le t\le T}|\varrho_{tN}| &\le C'\sup_{1\le t\le T}[N^{-1}\sum\nolimits_{i=1}^N\{\eta_{N,jt}(X_{ji})\}^2]^{1/2}[N^{-1}\sum\nolimits_{i=1}^N\{h_{jt}^0(X_{ji})\}^2]^{1/2}\\
&= O_p(K_N^{-r}) + O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}). \qquad \text{(A.22)}
\end{aligned}$$

Hence by (A.16), (A.21) and the fact that $f_{jt}^0 = \sqrt{\lim_{N\to\infty} N^{-1}\sum_{i=1}^N h_{jt}^0(X_{ji})^2}$, we have with probability approaching 1, as $N\to\infty$,

$$\begin{aligned}
&\widetilde{h}_{jt}(x_j)/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2} - h_{jt}^0(x_j)/f_{jt}^0\\
&= \{\widetilde{h}_{jt}(x_j) - h_{jt}^0(x_j)\}/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2} + h_{jt}^0(x_j)\{1/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2} - 1/f_{jt}^0\}\\
&= 1_{j+1}^\intercal\mathbb{B}(x)D_{Nt,1}/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2} + h_{jt}^0(x_j)\{1/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2} - 1/f_{jt}^0\}\\
&+ \eta_{N,jt}(x_j)/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2}
\end{aligned}$$

$$\begin{aligned}
&= 1_{j+1}^\intercal\mathbb{B}(x)D_{Nt,1}/f_{jt}^0 + \{1_{j+1}^\intercal\mathbb{B}(x)D_{Nt,1} + h_{jt}^0(x_j)\}\{1/\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widetilde{h}_{jt}(X_{ji})^2} - 1/f_{jt}^0\}\\
&+ C''\eta_{N,jt}(x_j)\\
&= 1_{j+1}^\intercal\mathbb{B}(x)D_{Nt,1}/f_{jt}^0 + \{1_{j+1}^\intercal\mathbb{B}(x)D_{Nt,1} + h_{jt}^0(x_j)\}C'N^{-1}\sum\nolimits_{i=1}^N 1_{j+1}^\intercal\mathbb{B}(x)D_{Nt,1}h_{jt}^0(X_{ji})\\
&+ C'''\varrho_{tN} + C''\eta_{N,jt}(x_j)
\end{aligned}$$

for some constants $0 < C'', C''' < \infty$.

Let $\varrho_N = T^{-1}\sum_{t=1}^T C'''\varrho_{tN}$ and $\eta_{NT,j}(x_j) = T^{-1}\sum_{t=1}^T C''\eta_{N,jt}(x_j)$. By (A.18) and (A.22), we have

$$|\varrho_N| = O_p(K_N^{-r}) + O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}), \qquad \text{(A.23)}$$

$$\begin{aligned}
\{\int \eta_{NT,j}(x_j)^2dx_j\}^{1/2} &= O_p(K_N^{-r}) + O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}),\\
\sup\nolimits_{x_j\in[a,b]}|\eta_{NT,j}(x_j)| &= O_p(K_N^{-r}) + O_p(K_N^{2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}). \qquad \text{(A.24)}
\end{aligned}$$

By the definitions of $\widehat{g}_j^{[0]}(x_j)$ and $g_j^0(x_j)$ given in (3.1) and (2.5), respectively, and $h_{jt}^0(X_{ji}) = g_j^0(X_{ji})f_{jt}^0$, we have with probability approaching 1, as $(N,T)\to\infty$,

$$\widehat{g}_j^{[0]}(x_j) - g_j^0(x_j) = \Phi_{NTj,1}(x_j) + \Phi_{NTj,2}(x_j) + \Phi_{NTj,3}(x_j) + \varrho_N + \eta_{NT,j}(x_j), \qquad \text{(A.25)}$$

where

$$\begin{aligned}
\Phi_{NTj,1}(x_j) &= T^{-1}\sum_{t=1}^{T} 1_{j+1}^{\intercal}\mathbb{B}(x)D_{Nt,1}/f_{jt}^0,\\
\Phi_{NTj,2}(x_j) &= C'(TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} 1_{j+1}^{\intercal}\mathbb{B}(X_i)D_{Nt,1}g_j^0(X_{ji})g_j^0(x_j)(f_{jt}^0)^2,\\
\Phi_{NTj,3}(x_j) &= C'(TN)^{-1}\sum_{t=1}^{T} 1_{j+1}^{\intercal}\mathbb{B}(x)D_{Nt,1}\sum_{i=1}^{N} 1_{j+1}^{\intercal}\mathbb{B}(X_i)D_{Nt,1}g_j^0(X_{ji})f_{jt}^0.
\end{aligned}$$

Define $\psi_{it,\ell} = Z_{i\ell}(\tau - I(\varepsilon_{it} < 0))(f_{jt}^0)^2$. Then $E(\psi_{it,\ell}) = 0$. Moreover, $E(\psi_{it,\ell})^2 \le c_1$ for some constant $0 < c_1 < \infty$, and by Condition (C1), we have

$$\begin{aligned}
|E(\psi_{it,\ell}\psi_{js,\ell})| &\le 2\{\phi(\sqrt{|i-j|^2+|t-s|^2})\}^{1/2}\{E(\psi_{it,\ell})^2E(\psi_{js,\ell})^2\}^{1/2}\\
&\le 2c_1\{\phi(\sqrt{|i-j|^2+|t-s|^2})\}^{1/2}.
\end{aligned}$$

Hence by Condition (C1), we have

$$\begin{aligned}
&E((NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\psi_{it,\ell})^2\\
&= (NT)^{-2}\sum_{t,t'}\sum_{i,i'}E(\psi_{it,\ell}\psi_{i't',\ell}) \le 2c_1(NT)^{-2}\sum_{t,t'}\sum_{i,i'}\{\phi(\sqrt{|i-j|^2+|t-s|^2})\}^{1/2}\\
&\le 2c_1K_1(NT)^{-2}\sum_{t,t'}\sum_{i,i'}e^{-\lambda_1\sqrt{|i-i'|^2+|t-t'|^2}/2}\\
&\le 2c_1(NT)^{-2}K_1\sum_{t,t'}\sum_{i,i'}e^{-(\lambda_1/2)(|i-i'|+|t-t'|)}\\
&\le 2c_1K_1(NT)^{-2}(NT)(\sum_{k=0}^{T}e^{-(\lambda_1/2)k})(\sum_{k=0}^{N}e^{-(\lambda_1/2)k})\\
&\le 2c_1K_1(NT)^{-2}(NT)\{1-e^{-(\lambda_1/2)}\}^{-2} = 2c_1K_1\{1-e^{-(\lambda_1/2)}\}^{-2}(NT)^{-1}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&E\left\|(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Z_i(\tau - I(\varepsilon_{it}<0))(f_{jt}^0)^2\right\|^2\\
&= \sum_{\ell=1}^{d(N)}E\{(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\psi_{it,\ell}\}^2 = O\{K_N(NT)^{-1}\}. \qquad \text{(A.26)}
\end{aligned}$$

Therefore, by Markov's inequality we have

$$\left\|(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Z_i(\tau - I(\varepsilon_{it}<0))(f_{jt}^0)^2\right\| = O_p[\{K_N(NT)^{-1}\}^{1/2}].$$

Moreover, $||N^{-1}\sum_{i=1}^{N}\mathbb{B}(X_i)^{\intercal}1_{j+1}g_j^0(X_{ji})|| = O_p(1)$ and $\sup_{x_j\in[a,b]}|g_j^0(x_j)| \le C'$ for some constant $C' \in (0,\infty)$ by Condition (C3). Hence by the above results and (A.9), we have

$$\begin{aligned}
\sup_{x_j\in[a,b]}|\Phi_{NTj,2}(x_j)| &\le C'\sup_{x_j\in[a,b]}|g_j^0(x_j)| \times ||N^{-1}\sum_{i=1}^{N}\mathbb{B}(X_i)^{\intercal}1_{j+1}g_j^0(X_{ji})|| \times ||\Psi_N^{-1}||\times\\
&\left\|(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Z_i(\tau - I(\varepsilon_{it}<0))(f_{jt}^0)^2\right\| = O_p\{\sqrt{K_N/(NT)}\}.
\end{aligned}$$

(A.27)

Moreover, by following the same procedure as the proof in (A.26), we have $E||N^{-1}\sum_{i=1}^N Z_i(\tau - I(\varepsilon_{it} < 0))||^2 = O_p(K_N N^{-1})$. Then we have $T^{-1}\sum_{t=1}^T ||N^{-1}\sum_{i=1}^N Z_i(\tau - I(\varepsilon_{it} < 0))||^2 = O_p(K_N N^{-1})$. Hence,

$$
\begin{aligned}
&\sup_{x_j\in[a,b]} |\Phi_{NTj,3}(x_j)| \\
&\leq T^{-1}\sum\nolimits_{t=1}^{T} \sup_{x\in[a,b]^J}\{1_{j+1}^{\intercal}\mathbb{B}(x)D_{Nt,1}\}^2 \sup_{x_j\in[a,b]}|g_j^0(x_j)||f_{jt}^0| \\
&\leq C\sup_{x\in[a,b]^J}||\mathbb{B}(x)^{\intercal}1_{j+1}||^2||\Psi_N^{-1}||^2 T^{-1}\sum\nolimits_{t=1}^{T}||N^{-1}\sum\nolimits_{i=1}^{N} Z_i(\tau - I(\varepsilon_{it}<0))||^2 \\
&= O_p(K_N^2 N^{-1}).
\end{aligned} \tag{A.28}
$$

By letting

$$\zeta_{NTj}(x_j) = \Phi_{NTj,2}(x_j) + \Phi_{NTj,3}(x_j) + \varrho_N + \eta_{NT,j}(x_j), \tag{A.29}$$

by (A.23), (A.24), (A.27) and (A.28), we have

$$
\begin{aligned}
\sup_{x_j\in[a,b]}|\zeta_{NTj}(x_j)| &= O_p(\sqrt{K_N/(NT)} + K_N^2 N^{-1} + K_N^2 N^{-3/4}\sqrt{\log NT} + K_N^{-r}) + o_p(N^{-1/2}) \\
&= O_p(\sqrt{K_N/(NT)} + K_N^2 N^{-3/4}\sqrt{\log NT} + K_N^{-r}) + o_p(N^{-1/2}), \\
\{\int \zeta_{NTj}(x_j)^2 dx_j\}^{1/2} &= O_p(\sqrt{K_N/(NT)} + K_N^{3/2} N^{-3/4}\sqrt{\log NT} + K_N^{-r}) + o_p(N^{-1/2}).
\end{aligned} \tag{A.30}
$$

Therefore, Proposition 1 follow from the above two results, (A.25) and (A.29). Moreover, by the definition of $D_{Nt,1}$ given in (A.11), we have

$$\Phi_{NTj,1}(x_j) = 1_{j+1}^{\intercal}\mathbb{B}(x)\Psi_N^{-1}\left[(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N} Z_i(\tau - I(\varepsilon_{it}<0))\right](f_{jt}^0)^{-1}.$$

Hence

$$
\begin{aligned}
\sup_{x_j\in[a,b]}|\Phi_{NTj,1}(x_j)| &\leq C_h^{-1}||\mathbb{B}(x)^{\intercal}1_{j+1}|| \times ||\Psi_N^{-1}|| \times ||(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N} Z_i(\tau - I(\varepsilon_{it}<0))|| \\
&= O_p\{K_N(NT)^{-1/2}\}
\end{aligned}
$$

$$
\begin{aligned}
&\{\int \Phi_{NTj,1}(x_j)^2 dx_j\}^{1/2} \leq C_h^{-1}[\lambda_{\max}[E\{B_j(X_{ji})B_j(X_{ji})^{\intercal}]]^{1/2}||\Psi_N^{-1}||\times \\
&||(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N} Z_i(\tau - I(\varepsilon_{it}<0))|| = O_p\{K_N^{1/2}(NT)^{-1/2}\}.
\end{aligned}
$$

Therefore, the result (A.1) follows from the above result, and (A.25), (A.29) and (A.30). □

8.2 Proofs of Theorems 1 and 2

We first present the following several lemmas that will be used in the proofs of Theorems 1 and 2. Lemmas 11-13 are used in the proof of Lemma 10, and Lemma 16 is used for the proof of Lemma 14. Lemmas 9, 10 and 14 are used in the proof of the main theorems. We define the infeasible estimator $f_t^* = \{f_{ut}^*, (f_{jt}^*, 1 \leq j \leq J)^{\intercal}\}^{\intercal}$ as the minimizer of

$$\sum\nolimits_{i=1}^{N}\rho_\tau(y_{it} - f_{ut} - \sum\nolimits_{j=1}^{J} g_j^0(X_{ji})f_{jt}). \tag{A.31}$$

Lemma 9. *Under Conditions (C1), (C2), (C4), (C5) and (C6), we have as* $N \to \infty$,

$$\sqrt{N}(\mathbf{\Sigma}_{Nt}^{0})^{-1/2}(f_t^* - f_t^0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}_{J+1}),$$

where $\mathbf{\Sigma}_{Nt}^{0}$ *is given in (4.4).*

Proof. By Bahadur representation for the ϕ-mixing case (see Babu (1989)), we have

$$f_t^* - f_t^0 = \Lambda_{Nt}^{-1}\{N^{-1}\sum\nolimits_{i=1}^{N} Q_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\} + \upsilon_{Nt}, \tag{A.32}$$

and $||\upsilon_{Nt}|| = o_p(N^{-1/2})$ for every t, where $\Lambda_{Nt} = N^{-1}\sum_{i=1}^{N} p_i\,(0\,|X_i, f_t)\,Q_i^0(X_i)Q_i^0(X_i)^\intercal$. By Conditions (C2), (C4) and (C5), we have that the eigenvalues of Λ_{Nt}^0 are bounded away from zero and infinity. By similar reasoning to the proof for Theorem 2 in Lee and Robinson (2016), we have $\|\Lambda_{Nt}^{-1}\| = O_p(1)$ and $\|\Lambda_{Nt} - \Lambda_{Nt}^0\| = o_p(1)$. Thus, the asymptotic distribution in Lemma 9 can be obtained directly by Condition (C6). □

Recall that the initial estimator $\widehat{f}_t^{[0]}$ given in (3.3) is defined in the same way as f_t^* with $g_j^0(X_{ji})$ replaced by $\widehat{g}_j^{[0]}(X_{ji})$ in (A.31). Then we have the following result for $\widehat{f}_t^{[0]}$.

Lemma 10. *Let Conditions (C1)-(C5) hold. If, in addition,* $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$ *and* $K_N^{-1}(\log NT)(\log N)^4 = o(1)$, *then for any* t *there is a stochastically bounded sequence* $\delta_{N,jt}$ *such that as* $N \to \infty$,

$$\sqrt{N}(\widehat{f}_t^{[0]} - f_t^* - d_{NT}\delta_{N,t}) = o_p(1),$$

where $\delta_{N,t} = (\delta_{N,jt}, 0 \le j \le J)^\intercal$ *and* d_{NT} *is given in (4.6).*

Proof. Denote $g = \{g_j(\cdot), 1 \le j \le J\}$. Define

$$\begin{aligned} L_{Nt}(f_t, g) = {} & N^{-1}\sum\nolimits_{i=1}^{N} \rho_\tau(y_{it} - f_{ut} - \sum\nolimits_{j=1}^{J} g_j(X_{ji})f_{jt}) \\ & - N^{-1}\sum\nolimits_{i=1}^{N} \rho_\tau(y_{it} - f_{ut}^0 - \sum\nolimits_{j=1}^{J} g_j(X_{ji})f_{jt}^0), \end{aligned}$$

so that f_t^* and $\widehat{f}_t^{[0]}$are the minimizers of $L_{Nt}(f_t, g^0)$ and $L_{Nt}(f_t, \widehat{g}^{[0]})$, respectively, where $\widehat{g}^{[0]} = \{\widehat{g}_j^{[0]}(\cdot), 1 \le j \le J\}$ and $g^0 = \{g_j^0(\cdot), 1 \le j \le J\}$. According to the result on page 149 of de Boor (2001), for g_j^0 satisfying the smoothness condition given in (C2), there exists $\boldsymbol{\lambda}_j^0 \in R^{K_n}$ such that $g_j^0(x_j) = \widetilde{g}_j^0(x_j) + r_j(x_j)$

$$\widetilde{g}_j^0(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j^0 \text{ and } \sup_j \sup\nolimits_{x_j \in [a,b]} |r_j(x_j)| = O(K_N^{-r}).$$

Since $\int\{\widehat{g}_j^{[0]}(x_j) - g_j^0(x_j)\}^2 dx_j = O_p(d_{NT}^2) + o_p(N^{-1/2})$ by Proposition 1, then there exists $\boldsymbol{\lambda}_{j,NT} \in R^{K_N}$ such that $\widehat{g}_j^{[0]}(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_{j,NT}$ and $||\boldsymbol{\lambda}_{j,NT} - \boldsymbol{\lambda}_j^0|| = O_p(d_{NT}^2) + o_p(N^{-1/2})$. In the following, we will show that for any $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$ not depending on f_t satisfying $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \le \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$ for some constant $0 < \widetilde{C} < \infty$, letting $\widetilde{f}_t$ be the minimizer of $L_{Nt}(f_t, g)$, we have

$$\widetilde{f}_t - f_t^0 - d_{NT}\delta_{N,t} = \Lambda_{Nt}^{-1}\{N^{-1}\sum\nolimits_{i=1}^{N} Q_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\} + o_p(N^{-1/2}). \tag{A.33}$$

Hence the result in Lemma 10 follows from (A.32) and (A.33). We have $||\widetilde{f}_t - f_t^*|| = o_p(1)$, since

$$
\begin{aligned}
&|L_{Nt}(f_t, g) - L_{Nt}(f_t, g^0)| \\
&\leq 2N^{-1}\sum\nolimits_{i=1}^{N} |\sum\nolimits_{j=1}^{J}\{g_j(X_{ji}) - g_j^0(X_{ji})\}f_{jt}| + 2N^{-1}\sum\nolimits_{i=1}^{N} |\sum\nolimits_{j=1}^{J}\{g_j(X_{ji}) - g_j^0(X_{ji})\}f_{jt}^0| \\
&\leq C_L\widetilde{C}\{d_{NT} + o(N^{-1/2})\} = o(1),
\end{aligned}
$$

for some constant $0 < C_L < \infty$, where the first inequality follows from the fact that $|\rho_\tau(u - v) - \rho_\tau(u)| \leq 2|v|$. Thus $||\widetilde{f}_t - f_t^0|| = o_p(1)$. Let $\mathbf{X} = (X_1, \ldots, X_N)^\intercal$, $Q_i(X_i) = \{1, g_1(X_{1i}), \ldots, g_J(X_{Ji})\}^\intercal$ and $\mathbf{F} = \{f_1, \ldots, f_T\}^\intercal$. For $g_j(x_j) = B_j(x_j)^\intercal\boldsymbol{\lambda}_j$ satisfying $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$ and f_t in a neighborhood of f_t^0, write

$$
\begin{aligned}
&L_{Nt}(f_t, g) = E\{L_{Nt}(f_t, g)|\mathbf{X}, \mathbf{F}\} - (f_t - f_t^0)^\intercal\{W_{Nt,1} - E(W_{Nt,1}|\mathbf{X}, \mathbf{F})\} \\
&+W_{Nt,2}(f_t, g) - E(W_{Nt,2}(f_t, g)|\mathbf{X}, \mathbf{F}),
\end{aligned} \tag{A.34}
$$

where

$$
W_{Nt,1} = N^{-1}\sum\nolimits_{i=1}^{N} Q_i(X_i)\psi_\tau(y_{it} - f_t^{0\intercal}Q_i(X_i)), \tag{A.35}
$$

$$
\begin{aligned}
W_{Nt,2}(f_t, g) = N^{-1}\sum\nolimits_{i=1}^{N}&\{\rho_\tau(y_{it} - f_t^\intercal Q_i(X_i)) - \rho_\tau(y_{it} - f_t^{0\intercal}Q_i(X_i)) \\
&+ (f_t - f_t^0)^\intercal Q_i(X_i)\psi_\tau(y_{it} - f_t^{0\intercal}Q_i(X_i))\}.
\end{aligned} \tag{A.36}
$$

In Lemma 11, we will show that as $N \to \infty$

$$
E\{L_{Nt}(f_t, g)|\mathbf{X}, \mathbf{F}\} = -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}, \mathbf{F}) + \frac{1}{2}(f_t - f_t^0)^\intercal\Lambda_{Nt}^0(f_t - f_t^0) + o_p(||f_t - f_t^0||^2),
$$

where $g_j(x_j) = B_j(x_j)^\intercal\boldsymbol{\lambda}_j$, uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$ and $||f_t - f_t^0|| \leq \varpi_N$, where ϖ_N is any sequence of positive numbers satisfying $\varpi_N = o(1)$. Substituting this into (A.34), we have with probability approaching 1,

$$
\begin{aligned}
L_{Nt}(f_t, g) = &-(f_t - f_t^0)^\intercal W_{Nt,1} + \frac{1}{2}(f_t - f_t^0)^\intercal\Lambda_N^0(f_t - f_t^0) \\
&+ W_{Nt,2}(f_t, g) - E(W_{Nt,2}(f_t, g)|\mathbf{X}, \mathbf{F}) + o(||f_t - f_t^0||^2).
\end{aligned}
$$

In Lemma 12, we will show that $W_{Nt,2}(f_t, g) - E(W_{Nt,2}(f_t, g)|\mathbf{X}, \mathbf{F}) = o_p(||f_t - f_t^0||^2 + N^{-1})$, where $g_j(x_j) = B_j(x_j)^\intercal\boldsymbol{\lambda}_j$, uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}d_{NT}$ and $||f_t - f_t^0|| \leq \varpi_N$. Thus, we have $\widetilde{f}_t - f_t^0 = (\Lambda_{Nt}^0)^{-1}W_{Nt,1} + o_p(N^{-1/2})$. Since $||(\Lambda_{Nt}^0)^{-1} - (\Lambda_{Nt})^{-1}|| = o_p(1)$, we have

$$
\widetilde{f}_t - f_t^0 = \Lambda_{Nt}^{-1}W_{Nt,1} + o_p(N^{-1/2}). \tag{A.37}
$$

In Lemma 13, we will show that for any t there is a stochastically bounded sequence $\delta_{N,jt}$ such that as $N \to \infty$,

$$
W_{Nt,1} = N^{-1}\sum\nolimits_{i=1}^{N} Q_i^0(X_i)\psi_\tau(\varepsilon_{it}) + d_{NT}\delta_{N,t} + o_p(N^{-1/2}). \tag{A.38}
$$

where $\delta_{N,t} = (\delta_{N,jt}, 0 \leq j \leq J)^\intercal$ and $g_j(x_j) = B_j(x_j)^\intercal\boldsymbol{\lambda}_j$, uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$. Hence, result (A.33) follows from (A.37) and (A.38) directly. Then the proof is complete. □

Lemma 11. *Under Conditions (C2), (C4) and (C5),*

$$E\{L_{Nt}(f_t, g)|\mathbf{X}, \mathbf{F}\} = -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}, \mathbf{F}) + \frac{1}{2}(f_t - f_t^0)^\intercal \Lambda_{Nt}^0 (f_t - f_t^0) + o_p(||f_t - f_t^0||^2),$$

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$ *and* $||f_t - f_t^0|| \leq \varpi_N$*, where* $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$ *and* ϖ_N *is any sequence of positive numbers satisfying* $\varpi_N = o(1)$*.*

Proof. By using the identity of Knight (1998) that

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (I(u \leq s) - I(u \leq 0))ds, \qquad \text{(A.39)}$$

we have

$$\begin{aligned} &\rho_\tau(y_{it} - f_t^\intercal Q_i(X_i)) - \rho_\tau(y_{it} - f_t^{0\intercal} Q_i(X_i)) \\ &= -(f_t - f_t^0)^\intercal Q_i(X_i)\psi_\tau(y_{it} - f_t^{0\intercal} Q_i(X_i)) \\ &+ \int_0^{(f_t - f_t^0)^\intercal Q_i(X_i)} \Big(I(y_{it} - f_t^{0\intercal} Q_i(X_i) \leq s) - I(y_{it} - f_t^{0\intercal} Q_i(X_i) \leq 0)\Big)\, ds. \end{aligned} \qquad \text{(A.40)}$$

By Lipschitz continuity of $p_i(\varepsilon|X_i, f_t)$ given in Condition (C1) and boundedness of f_{jt}^0 in Condition (C3), we have

$$\begin{aligned} &F_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i)) + s|X_i, f_t\} - F_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i))|X_i, f_t\} \\ &= sp_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i))|X_i, f_t\} + o(s), \end{aligned}$$

where $o(\cdot)$ holds uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$ and $||f_t - f_t^0|| \leq \varpi_N$. Then we have

$$\begin{aligned} &E\{L_{Nt}(f_t, g)|\mathbf{X}, \mathbf{F}\} \\ &= -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}, \mathbf{F}) + N^{-1}\sum\nolimits_{i=1}^N \int_0^{(f_t - f_t^0)^\intercal Q_i(X_i)} [F_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i)) + s|X_i, f_t\} \\ &- F_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i))|X_i, f_t\}]ds \\ &= -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}, \mathbf{F}) + N^{-1}\sum\nolimits_{i=1}^N \int_0^{(f_t - f_t^0)^\intercal Q_i(X_i)} [sp_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i))|X_i, f_t\}]ds \\ &+ o\left[(f_t - f_t^0)^\intercal\{N^{-1}\sum\nolimits_{i=1}^N Q_i(X_i)Q_i(X_i)^\intercal\}(f_t - f_t^0)\right] \\ &= -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}, \mathbf{F}) + \frac{1}{2}(f_t - f_t^0)^\intercal \times \\ &\left[N^{-1}\sum\nolimits_{i=1}^N p_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i))|X_i, f_t\}Q_i(X_i)Q_i(X_i)^\intercal\right](f_t - f_t^0) \\ &+ o\left[(f_t - f_t^0)^\intercal\{N^{-1}\sum\nolimits_{i=1}^N Q_i(X_i)Q_i(X_i)^\intercal\}(f_t - f_t^0)\right]. \end{aligned} \qquad \text{(A.41)}$$

Since $\sup_{x_j \in [a,b]} |g_j(x_j) - g_j^0(x_j)| = o(1)$, then $\sup_{x \in \mathcal{X}} |f_t^{0\intercal}(Q_i(x) - Q_i^0(x))| = o(1)$. By similar reasoning to the proof for Theorem 2 in Lee and Robinson (2016), we have $N^{-1}\sum_{i=1}^N Q_i(X_i)Q_i(X_i)^\intercal = E\{Q_i(X_i)Q_i(X_i)^\intercal\} + o_p(1)$. Hence, by these results and Condition (C4), we have the result in Lemma 11. □

Lemma 12. *Under Conditions (C2), (C4) and (C5), we have*

$$W_{Nt,2}(f_t,g) - E(W_{Nt,2}(f_t,g)|\mathbf{X},\mathbf{F}) = o_p(||f_t - f_t^0||^2 + N^{-1})$$

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \le \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$ *and* $||f_t - f_t^0|| \le \varpi_N$*, where* $W_{Nt,2}(f_t,g)$ *is defined in (A.36),* $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$*, and* ϖ_N *is any sequence of positive numbers satisfying* $\varpi_N = o(1)$*.*

Proof. By (A.40), we have

$$W_{Nt,2i}(f_t,g) = \int_0^{(f_t-f_t^0)^\intercal Q_i(X_i)} \Big(I(y_{it} - f_t^{0\intercal} Q_i(X_i) \le s) - I(y_{it} - f_t^{0\intercal} Q_i(X_i) \le 0)\Big)\, ds,$$

and thus

$$\begin{aligned} E(W_{Nt,2i}(f_t,g)|X_i,f_t) = &\int_0^{(f_t-f_t^0)^\intercal Q_i(X_i)} [F_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i)) + s|X_i,f_t\} \\ &- F_i\{f_t^{0\intercal}(Q_i(X_i) - Q_i^0(X_i))|X_i,f_t\}]ds. \end{aligned}$$

By following the same reasoning as the proof for (A.41), we have

$$\sup_{X_i \in [a,b]^J} |E(W_{Nt,2i}(f_t,g)|X_i,f_t) - \frac{1}{2}(f_t - f_t^0)^\intercal p_i(0|X_i,f_t) Q_i(X_i) Q_i(X_i)^\intercal (f_t - f_t^0)| = o_p(||f_t - f_t^0||^2).$$

Hence with probability approaching 1, as $N \to \infty$,

$$\sup_{X_i \in [a,b]^J} |E(W_{Nt,2i}(f_t,g)|X_i,f_t)| \le C_W ||f_t - f_t^0||^2,$$

for some constant $0 < C_W < \infty$. Moreover,

$$\begin{aligned} &E\{W_{Nt,2i}(f_t,g)\}^2 \\ &= E[E[\{\int_0^{(f_t-f_t^0)^\intercal Q_i(X_i)} (I(y_{it} - f_t^{0\intercal}Q_i(X_i) \le s) - I(y_{it} - f_t^{0\intercal}Q_i(X_i) \le 0))ds\}^2 |X_i,f_t]] \\ &\le E[E[|I(y_{it} - f_t^{0\intercal}Q_i(X_i) \le (f_t - f_t^0)^\intercal Q_i(X_i)) - I(y_{it} - f_t^{0\intercal}Q_i(X_i) \le 0)| \\ &\times \{(f_t - f_t^0)^\intercal Q_i(X_i)\}^2 |X_i,f_t]] \\ &= E[E[|I(\varepsilon_{it} \le f_t^\intercal Q_i(X_i) - f_t^{0\intercal}Q_i(X_i)^0) - I(\varepsilon_{it} \le f_t^{0\intercal}(Q_i(X_i) - Q_i(X_i)^0)| \\ &\times \{(f_t - f_t^0)^\intercal Q_i(X_i)\}^2 |X_i,f_t]] \\ &\le C'' E|(f_t - f_t^0)^\intercal Q_i(X_i)|^3 \le C''' E||f_t - f_t^0||^3 \end{aligned} \tag{A.42}$$

for some constants $0 < C'' < \infty$ and $0 < C''' < \infty$. Therefore, for $N \to \infty$,

$$\begin{aligned} &E\{W_{Nt,2}(f_t,g) - E(W_{Nt,2}(f_t,g)|\mathbf{X},\mathbf{F})\}^2 \\ &= N^{-2} \sum\nolimits_{i=1}^N E\left[W_{Nt,2i}(f_t,g) - E(W_{Nt,2i}(f_t,g)|X_i,f_t)\right]^2 \\ &\le N^{-2} \sum\nolimits_{i=1}^N [2E\{W_{Nt,2i}(f_t,g)\}^2 + 2E[E(W_{Nt,2i}(f_t,g)|X_i,f_t)]^2] \\ &\le N^{-1}(2C''' E||f_t - f_t^0||^3 + 2C_W^2 E||f_t - f_t^0||^4) \le C'''' N^{-1} E||f_t - f_t^0||^3, \end{aligned}$$

for some constant $0 < C'''' < \infty$. By following the same routine procedure as the proof in Lemma 5 by appyling the Bernstein's inequality, we have

$$\sup_{||\boldsymbol{\lambda}_j-\boldsymbol{\lambda}_j^0||\leq \widetilde{C}d_{NT}, ||f_t-f_t^0||\leq \varpi_N} ||f_t - f_t^0||^{-3/2}|W_{Nt,2}(f_t, g) - E(W_{Nt,2}(f_t, g)|\mathbf{X}, \mathbf{F})| = O_p(N^{-1/2}).$$

Hence, we have $|W_{Nt,2}(f_t, g) - E(W_{Nt,2}(f_t, g)|\mathbf{X}, \mathbf{F})| = O_p(||f_t - f_t^0||^{-3/2}N^{-1/2})$, uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}d_{NT}$ and $||f_t - f_t^0|| \leq \varpi_N$. Since

$$\begin{aligned}
N^{-1/2}||f_t - f_t^0||^{3/2} &\leq N^{-1}||f_t - f_t^0||^{1/2} + ||f_t - f_t^0||^2||f_t - f_t^0||^{1/2} \\
&\leq N^{-1}\varpi_N + ||f_t - f_t^0||^2\varpi_N,
\end{aligned}$$

then we have $W_{Nt,2}(f_t, g) - E(W_{Nt,2}(f_t, g)|\mathbf{X}, \mathbf{F}) = o_p(||f_t - f_t^0||^2 + N^{-1})$, uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}d_{NT}$ and $||f_t - f_t^0|| \leq \varpi_N$. □

Lemma 13. *Under Conditions (C1), (C2), (C4) and (C5), for any t there is a stochastically bounded sequence $\delta_{N,jt}$ such that as $N \to \infty$,*

$$W_{Nt,1} = N^{-1}\sum\nolimits_{i=1}^{N} G_i^0(X_i)\psi_\tau(\varepsilon_{it}) + d_{NT}\delta_{N,t} + o_p(N^{-1/2}),$$

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}\{d_{NT} + o(N^{-1/2})\}$, where $W_{Nt,1}$ is defined in (A.35), $\delta_{N,t} = (\delta_{N,jt}, 0 \leq j \leq J)^\intercal$ and $g_j(x_j) = B_j(x_j)^\intercal\boldsymbol{\lambda}_j$.

Proof. Write

$$W_{Nt,1} = W_{Nt,11} + W_{Nt,12} + W_{Nt,13}, \tag{A.43}$$

where

$$\begin{aligned}
W_{Nt,11} &= N^{-1}\sum\nolimits_{i=1}^{N} Q_i^0(X_i)\psi_\tau(y_{it} - f_t^{0\intercal}Q_i^0(X_i)), \\
W_{Nt,12} &= (W_{Ntj,12}, 0 \leq j \leq J)^\intercal = N^{-1}\sum\nolimits_{i=1}^{N}(Q_i(X_i) - Q_i^0(X_i))\psi_\tau(y_{it} - f_t^{0\intercal}Q_i^0(X_i)), \\
W_{Nt,13} &= (W_{Ntj,13}, 0 \leq j \leq J)^\intercal \\
&= N^{-1}\sum\nolimits_{i=1}^{N} Q_i(X_i)\{\psi_\tau(y_{it} - f_t^{0\intercal}Q_i(X_i)) - \psi_\tau(y_{it} - f_t^{0\intercal}Q_i^0(X_i))\}.
\end{aligned}$$

It is easy to see that $E(W_{Ntj,12}) = 0$. Also by the ϕ-mixing distribution condition given in Condition (C1), we have $\text{var}(W_{Ntj,12}) \leq C_{W_{12}}N^{-1}d_{NT}^2$ for some constant $0 < C_{W_{12}} < \infty$, then by following the routine procedure as the proof in Lemma 5, we have

$$\sup\nolimits_{||\boldsymbol{\lambda}_j-\boldsymbol{\lambda}_j^0||\leq \widetilde{C}d_{NT}} |W_{Ntj,12}| = o_p(N^{-1/2}). \tag{A.44}$$

Moreover,

$$\begin{aligned}
E(W_{Ntj,13}|\mathbf{X}, \mathbf{F}) &= N^{-1}\sum\nolimits_{i=1}^{N} g_j(X_{ji})E\{I(y_{it} - f_t^{0\intercal}Q_i^0(X_i) \leq 0) - I(y_{it} - f_t^{0\intercal}Q_i(X_i) \leq 0)|X_i, f_t\} \\
&= N^{-1}\sum\nolimits_{i=1}^{N} g_j(X_{ji})\int_{f_t^{0\intercal}(Q_i(X_i)-Q_i^0(X_i))}^{0} p_i(s|X_i, f_t)ds \\
&= N^{-1}\sum\nolimits_{i=1}^{N} g_j(X_{ji})p_i(0|X_i, f_t)f_t^{0\intercal}(Q_i^0(X_i) - Q_i(X_i)) + O(d_{NT}^2) + o(N^{-1}).
\end{aligned}$$

Let

$$d_{NT}\delta_{N,jt} = N^{-1}\sum\nolimits_{i=1}^{N} g_j(X_{ji})p_i(0|X_i, f_t)f_t^{0\intercal}(Q_i^0(X_i) - Q_i(X_i)) + O(d_{NT}^2).$$

Since $N^{-1}\sum_{i=1}^{N}\{g_j(X_{ji}) - g_j^0(X_{ji})\}^2 \leq \{\widetilde{C}(d_{NT} + o(N^{-1/2}))\}^2$, then as $N \to \infty$, $|d_{NT}\delta_{N,jt}| \leq C_\delta(d_{NT} + o(N^{-1/2}))$ for some constant $0 < C_\delta < \infty$. Therefore,

$$E(W_{Ntj,13}|\mathbf{X}, \mathbf{F}) = d_{NT}\delta_{N,jt} + o(N^{-1/2}). \tag{A.45}$$

Also by the ϕ–mixing condition given in Condition (C1), we have $E\{W_{Ntj,13} - E(W_{Ntj,13}|\mathbf{X}, \mathbf{F})\}^2 \leq C'_\delta N^{-1}d_{NT}$ for some constant $0 < C'_\delta < \infty$. Therefore, by following the procedure as the proof in Lemma 5, we have

$$\sup\nolimits_{||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C}d_{NT}} |W_{Ntj,13} - E(W_{Ntj,13}|\mathbf{X}, \mathbf{F})| = o_p(N^{-1/2}). \tag{A.46}$$

Therefore, the result in Lemma 13 is proved by (A.43), (A.44), (A.45) and (A.46). □

Let $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\intercal, \ldots, \boldsymbol{\lambda}_J^\intercal)^\intercal$. For given $\widehat{f}^{[0]}$, we obtain

$$\widehat{\boldsymbol{\lambda}}^{[1]} = \arg\min_{\boldsymbol{\lambda}}\{(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\rho_\tau(y_{it} - \widehat{f}_{ut}^{[0]} - \sum\nolimits_{j=1}^{J} B_j(X_{ji})^\intercal\boldsymbol{\lambda}_j\widehat{f}_{jt}^{[0]})\}.$$

Let $\widehat{g}_j^{*[1]}(x_j) = B_j(x_j)^\intercal\widehat{\boldsymbol{\lambda}}_j^{[1]}$. The estimate for $g_j(x_j)$ at the 1st step is

$$\widehat{g}_j^{[1]}(x_j) = \widehat{g}_j^{*[1]}(x_j)/\sqrt{N^{-1}\sum\nolimits_{i=1}^{N}\widehat{g}_j^{*[1]}(X_{ji})^2}.$$

We define the infeasible estimator of $\boldsymbol{\lambda}$ as

$$\boldsymbol{\lambda}^* = \arg\min_{\boldsymbol{\lambda}}\{(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\rho_\tau(y_{it} - f_{ut}^0 - \sum\nolimits_{j=1}^{J} B_j(X_{ji})^\intercal\boldsymbol{\lambda}_j f_{jt}^0)\}.$$

Let $g_j^*(x_j) = B_j(x_j)^\intercal\boldsymbol{\lambda}_j^*$ and $\widetilde{g}_j^*(x_j) = g_j^*(x_j)/\sqrt{N^{-1}\sum_{i=1}^{N} g_j^*(X_{ji})^2}$.

Lemma 14. *Let Conditions (C1)–(C5) hold. If, in addition,* $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$ *and* $K_N^{-1}(\log NT)(\log N)^4 = o(1)$, *then for every* $1 \leq j \leq J$,

$$\left[\int\{\widetilde{g}_j^*(x_j) - g_j^0(x_j)\}^2 dx_j\right]^{1/2} = O_p(K_N^{1/2}(NT)^{-1/2} + K_N^{-r}), \tag{A.47}$$

and

$$\int\{\widehat{g}_j^{[1]}(x_j)(x_j) - \widetilde{g}_j^*(x_j)\}^2 dx_j = O_p(d_{NT}^2) + o_p(N^{-1/2}). \tag{A.48}$$

Therefore, for every $1 \leq j \leq J$,

$$\int\{\widehat{g}_j^{[1]}(x_j) - g_j^0(x_j)\}^2 dx_j = O_p(d_{NT}^2) + o_p(N^{-1/2}). \tag{A.49}$$

Proof. Denote $\widetilde{g}^0(x) = \{\widetilde{g}_j^0(x_j), 1 \leq j \leq J\}^\intercal$ and $g^*(x) = \{g_j^*(x_j), 1 \leq j \leq J\}^\intercal$. Let $\boldsymbol{\lambda}^0 = (\boldsymbol{\lambda}_1^{0\intercal}, \ldots, \boldsymbol{\lambda}_J^{0\intercal})^\intercal$. Let $\mathbb{B}^*(x) = [\mathrm{diag}\,[B_1(x_1)^\intercal, \ldots, B_J(x_J)^\intercal]]_{J\times JK_N}$. Then $\mathbb{B}^*(x)\boldsymbol{\lambda}^* = g^*(x)$ and $\mathbb{B}^*(x)\boldsymbol{\lambda}^0 = \widetilde{g}^0(x)$. Let $W_{it}^0 = \{B_j(X_{ji})^\intercal f_{jt}^0, 1 \leq j \leq J\}^\intercal$,

$$\Psi_{NT} = (NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} f_\varepsilon\left(0\,|X_i, f_t\right) W_{it}^0 W_{it}^{0\intercal},$$

and $r^*_{j,it} = r_j(X_{ji})f^0_{jt}$. Moreover, define

$$U_{NT,1} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} W^0_{it}(\tau - I(\varepsilon_{it} < 0)), \tag{A.50}$$

$$U_{NT,2} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} W^0_{it} f_\varepsilon(0|X_i, f_t)\left(\sum_{j=1}^{J} r^*_{j,it}\right). \tag{A.51}$$

Let 1_j be the $J \times 1$ vector with the j^{th} component being one and others being zero. By the same procedure as the proof of Lemma 8, for $K_N^4(\log(NT))^2(NT)^{-1} = o(1)$, we obtain the Bahadur representation for $\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^0$ as

$$\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^0 = \Psi_{NT}^{-1}(U_{N,1} + U_{N,2}) + R^*_{NT} \tag{A.52}$$

and the remaining term R^*_{NT} satisfies

$$\begin{aligned}||R^*_{NT}|| &= O_p(K_N^{3/2}(NT)^{-1} + K_N^{3/2}(NT)^{-3/4}\sqrt{\log(NT)} + K_N^{1/2-2r} + (NT)^{-1/2}K_N^{-r/2+1/2}) \\ &= O_p(K_N^{3/2}(NT)^{-3/4}\sqrt{\log(NT)} + K_N^{1/2-2r}) + o_p((NT)^{-1/2}).\end{aligned}$$

By (A.52) and following the same reasoning as the proof for (A.20), we have $\sup_{x_j\in[a,b]}|g^*_j(x_j) - g^0_j(x_j)| = O_p(K_N(NT)^{-1/2} + K_N^{-r})$, $[\int\{g^*_j(x_j) - g^0_j(x_j)\}^2dx_j]^{1/2} = O_p(K_N^{1/2}(NT)^{-1/2} + K_N^{-r})$, and $[N^{-1}\sum_{i=1}^{N}\{g^*_j(X_{ji}) - g^0_j(X_{ji})\}^2]^{1/2} = O_p(K_N^{1/2}(NT)^{-1/2} + K_N^{-r})$. Therefore, we have

$$\{\sqrt{N^{-1}\sum_{i=1}^{N} g^*_j(X_{ji})^2}\}^{-1} - \{\sqrt{N^{-1}\sum_{i=1}^{N} g^0_j(X_{ji})^2}\}^{-1} = O_p(K_N^{1/2}(NT)^{-1/2} + K_N^{-r}),$$

and thus

$$\sup_{x_j\in[a,b]}|\widetilde{g}^*_j(x_j) - g^0_j(x_j)| = O_p(K_N(NT)^{-1/2} + K_N^{-r}),$$

$$[\int\{\widetilde{g}^*_j(x_j) - g^0_j(x_j)\}^2dx_j]^{1/2} = O_p(K_N^{1/2}(NT)^{-1/2} + K_N^{-r}).$$

Then the result (A.47) is proved. Define

$$\begin{aligned}L^*_{NT}(f,\boldsymbol{\lambda}) = (NT)^{-1}&\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(y_{it} - f_{ut} - \sum_{j=1}^{J} B_j(X_{ji})^\intercal\boldsymbol{\lambda}_j f_{jt}) \\ &- (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(y_{it} - f_{ut} - \sum_{j=1}^{J} B_j(X_{ji})^\intercal\boldsymbol{\lambda}^0_j f_{jt}).\end{aligned}$$

Hence, $\widehat{\boldsymbol{\lambda}}^{[1]}$ and $\boldsymbol{\lambda}^*$ are the minimizers of $L^*_{NT}(\widehat{f}^{[0]},\boldsymbol{\lambda})$ and $L^*_{NT}(f^0,\boldsymbol{\lambda})$, respectively. In Lemma 15, we will show that

$$||\widehat{\boldsymbol{\lambda}}^{[1]} - \boldsymbol{\lambda}^0 - \Psi_{NT}^{-1}U_{N,1}|| = O_p(d_{NT}) + o_p(N^{-1/2}). \tag{A.53}$$

Hence, by (A.52), (A.53) and $||\Psi_{NT}^{-1}U_{N,2}|| = O(K_N^{-r})$, we have

$$||\widehat{\boldsymbol{\lambda}}^{[1]} - \boldsymbol{\lambda}^*|| = O_p(d_{NT}) + o_p(N^{-1/2}). \tag{A.54}$$

Then we have $\int\{\widehat{g}^{*[1]}_j(x_j) - g^*_j(x_j)\}^2dx_j = O_p(d^2_{NT})$ and $N^{-1}\sum_{i=1}^{N}\{\widehat{g}^{*[1]}_j(x_j) - g^*_j(X_{ji})\}^2 = O_p(d^2_{NT})$. Thus,

$$\{\sqrt{N^{-1}\sum_{i=1}^{N}\widehat{g}^{*[1]}_j(X_{ji})^2}\}^{-1} - \{\sqrt{N^{-1}\sum_{i=1}^{N} g^*_j(X_{ji})^2}\}^{-1} = O_p(K_N^{1/2}(NT)^{-1/2} + K_N^{-r}),$$

and the result (A.48) follows from the above results directly. □

Lemma 15. *Let Conditions (C1)-(C4) hold. If, in addition, $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$ and $K_N^{-1}(\log NT)(\log N)^4 = o(1)$, then we have*

$$||\widehat{\boldsymbol{\lambda}}^{[1]} - \boldsymbol{\lambda}^0 - \Psi_{NT}^{-1} U_{N,1}|| = O_p(d_{NT}) + o_p(N^{-1/2}),$$

where $U_{NT,1}$ is defined in (A.50).

Proof. By Lemma A.32 and (A.32), we have $||\widehat{f}_t^{[0]} - f_t^0|| \leq C_f(d_{NT} + N^{-1/2})$ for some constant $0 < C_f < \infty$. Let $W_{it} = \{B_j(X_{ji})^\intercal f_{jt}, 1 \leq j \leq J\}^\intercal$. Let $f = (f_1^\intercal, \dots, f_T^\intercal)^\intercal$ satisfying that $||f_t - f_t^0|| \leq C_f(d_{NT} + N^{-1/2})$. Write

$$\begin{aligned}
&L_{NT}^*(f, \boldsymbol{\lambda}) \\
&= E\{L_{NT}^*(f, \boldsymbol{\lambda})|\mathbf{X}, \mathbf{F}\} - (\boldsymbol{\lambda} - \boldsymbol{\lambda}^0)^\intercal \{V_{NT,1}(f) - E(V_{NT,1}(f)|\mathbf{X}, \mathbf{F})\} \\
&+ V_{NT,2}(f, \boldsymbol{\lambda}) - E(V_{NT,2}(f, \boldsymbol{\lambda})|\mathbf{X}, \mathbf{F}), \qquad \text{(A.55)}
\end{aligned}$$

where

$$\begin{aligned}
V_{NT,1}(f) &= (NT)^{-1} \sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T W_{it} \psi_\tau(y_{it} - f_{ut} - \boldsymbol{\lambda}^{0\intercal} W_{it}), \qquad \text{(A.56)} \\
V_{NT,2}(f, \boldsymbol{\lambda}) &= (NT)^{-1} \sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T \{\rho_\tau(y_{it} - f_{ut} - \boldsymbol{\lambda}^\intercal W_{it}) - \rho_\tau(y_{it} - f_{ut} - \boldsymbol{\lambda}^{0\intercal} W_{it}) \\
&\quad + (\boldsymbol{\lambda} - \boldsymbol{\lambda}^0)^\intercal W_{it} \psi_\tau(y_{it} - f_{ut} - \boldsymbol{\lambda}^{0\intercal} W_{it})\}.
\end{aligned}$$

By following the same reasoning as in the proofs of Lemmas 11 and 12, we have

$$E\{L_{NT}^*(f, \boldsymbol{\lambda})|\mathbf{X}\} = -(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0)^\intercal E(V_{NT,1}(f)|\mathbf{X}, \mathbf{F}) + \frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0)^\intercal \Psi_{NT}(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0) + o_p(||\boldsymbol{\lambda}-\boldsymbol{\lambda}^0||^2), \quad \text{(A.57)}$$

$$V_{NT,2}(f, \boldsymbol{\lambda}) - E(V_{NT,2}(f, \boldsymbol{\lambda})|\mathbf{X}, \mathbf{F}) = o_p(||\boldsymbol{\lambda} - \boldsymbol{\lambda}^0||^2 + (NT)^{-1}), \qquad \text{(A.58)}$$

uniformly in $||f_t - f_t^0|| \leq C_f(d_{NT} + N^{-1/2})$ and $||\boldsymbol{\lambda} - \boldsymbol{\lambda}^0|| \leq \varsigma_{NT}$, where ς_{NT} is any sequence of positive numbers satisfying $\varsigma_{NT} = o(1)$.Thus, by (A.55), (A.57) and (A.58), we have

$$L_{NT}^*(f, \boldsymbol{\lambda}) = -(\boldsymbol{\lambda} - \boldsymbol{\lambda}^0)^\intercal V_{NT,1}(f) + \frac{1}{2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^0)^\intercal \Psi_{NT}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^0) + o_p(||\boldsymbol{\lambda} - \boldsymbol{\lambda}^0||^2 + (NT)^{-1}),$$

uniformly in $||f_t - f_t^0|| \leq C_f(d_{NT} + N^{-1/2})$ and $||\boldsymbol{\lambda} - \boldsymbol{\lambda}^0|| \leq \varsigma_{NT}$. Therefore, we have

$$\widehat{\boldsymbol{\lambda}}^{[1]} - \boldsymbol{\lambda}^0 = \Psi_{NT}^{-1} V_{NT,1}(\widehat{f}^{[0]}) + o_p\{(NT)^{-1/2}\},$$

By following the same reasoning as the proof for (A.9), as $(N, T) \to \infty$ with probability approaching 1, we have $||\Psi_{NT}^{-1}|| \leq C'_\Psi$ for some constant $0 < C'_\Psi < \infty$. In Lemma 16, we will show that $||V_{NT,1}(\widehat{f}^{[0]}) - U_{NT,1}|| = O_p(d_{NT}) + o_p(N^{-1/2})$. Therefore, the result in Lemma 15 follows from the above results, and thus the proof is completed. □

Lemma 16. *Let Conditions (C1)-(C4) hold. If, in addition, $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$ and $K_N^{-1}(\log NT)(\log N)^4 = o(1)$, then we have*

$$||V_{NT,1}(\widehat{f}^{[0]}) - U_{NT,1}|| = O_p(d_{NT}) + o_p(N^{-1/2}),$$

where $V_{NT,1}$ and $U_{NT,1}$ are defined in (A.56) and (A.50), respectively.

Proof. Write

$$V_{NT,1}(f) = V_{NT,11} + V_{NT,12}(f) + V_{NT,13}(f), \tag{A.59}$$

where

$$\begin{aligned} V_{NT,11} = U_{NT,1} &= (NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} W_{it}^{0}\psi_\tau(\varepsilon_{it}), \\ V_{NT,12}(f) &= (NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}(W_{it} - W_{it}^{0})\psi_\tau(\varepsilon_{it})), \\ V_{NT,13}(f) &= (NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} W_{it}\{\psi_\tau(y_{it} - f_{ut} - \boldsymbol{\lambda}^{0\intercal}W_{it})) - \psi_\tau(\varepsilon_{it})\}. \end{aligned}$$

Since $||N^{-1}\sum_{i=1}^{N} B(X_i)\psi_\tau(\varepsilon_{it})|| = O_p(N^{-1/2})$, we have with probability approaching 1,

$$\begin{aligned} &\sup_{||f_t - f_t^0|| \le C_f(d_{NT}+N^{-1/2})} ||V_{NT,12}|| \le T^{-1}\sum\nolimits_{t=1}^{T} ||N^{-1}\sum\nolimits_{i=1}^{N} B(X_i)\psi_\tau(\varepsilon_{it})|| \\ \times &\sup_{||f_t - f_t^0|| \le C_f(d_{NT}+N^{-1/2})} ||f_t - f_t^0|| = O\{N^{-1/2}(d_{NT} + N^{-1/2})\} = o(N^{-1/2} + d_{NT}). \end{aligned} \tag{A.60}$$

By following the same procedure as the proof for (A.68), we have for any vector $\mathbf{a} \in R^{K_N J}$ with $||\mathbf{a}|| = 1$,

$$\text{var}(\mathbf{a}^\intercal V_{NT,13}(f)\mathbf{a}) = O\{K_N(d_{NT} + N^{-1/2})(NT)^{-1}\},$$

uniformly in $||f_t - f_t^0|| \le C_f(d_{NT} + N^{-1/2})$. Then by the procedure as the proof in Lemma 5, we have

$$\begin{aligned} \sup_{||f_t - f_t^0|| \le C_f(d_{NT}+N^{-1/2})} ||V_{NT,13}(f) - E\{V_{NT,13}(f)\}|| &= O_p\{K_N^{1/2}(d_{NT} + N^{-1/2})^{1/2}(NT)^{-1/2}\} \\ &= o_p(d_{NT}). \end{aligned}$$

Hence,

$$||V_{NT,13}(\widehat{f}^{[0]}) - E\{V_{NT,13}(\widehat{f}^{[0]})\}|| = o_p(d_{NT}). \tag{A.61}$$

Let

$$\kappa_{it}(f) = f_{ut}^{0} - f_{ut} + \sum\nolimits_{j=1}^{J}(\widetilde{g}_j^{0}(X_{ji})(f_{jt}^{0} - f_{jt}) + r_{j,it}^{*}).$$

Then there exist constants $0 < C, C' < \infty$ such that

$$\begin{aligned} ||E\{V_{NT,13}(f)|\mathbf{X},\mathbf{F}\}|| &\le C||E[(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)\{I(\varepsilon_{it} \le 0) - I(\varepsilon_{it} \le \kappa_{it}(f))\}|\mathbf{X},\mathbf{F}]|| \\ &\le C'||(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)\kappa_{it}(f)p_i(0|X_i, f_t)|| \end{aligned} \tag{A.62}$$

uniformly in $||f_t - f_t^0|| \le C_f(d_{NT} + N^{-1/2})$. Moreover, by (A.32) and Lemma 10, we have

$$\begin{aligned} &\Bigg\|(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)\kappa_{it}(\widehat{f}^{[0]})p_i(0|X_i, f_t) \\ &+(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)p_i(0|X_i, f_t)\widetilde{g}^{0}(X_i)^\intercal[\Lambda_N^{-1}\{N^{-1}\sum\nolimits_{i=1}^{N} Q_i^{0}(X_i)(\tau - I(\varepsilon_{it} < 0))\}]\Bigg\| \\ &= O(d_{NT}) + o_p(N^{-1/2}). \end{aligned} \tag{A.63}$$

Since $||(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} Q_i^{0}(X_i)(\tau - I(\varepsilon_{it} < 0))|| = O_p\{(NT)^{-1/2}\}$, and

$$||(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)p_i(0|X_i, f_t)|| = O_p(1),$$

we have

$$\left\|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}B_i(X_i)p_i(0|X_i,f_t)\widehat{g}^0(X_i)^\intercal[\Lambda_N^{-1}\{N^{-1}\sum_{i=1}^{N}Q_i^0(X_i)(\tau-I(\varepsilon_{it}<0))\}]\right\|$$
$$=O_p\{(NT)^{-1/2}\}. \tag{A.64}$$

Therefore, by (A.62), (A.63), we have with probability approaching 1,

$$||E\{V_{NT,13}(\widehat{f}^{[0]})|\mathbf{X},\mathbf{F}\}||=O(d_{NT})+o(N^{-1/2}). \tag{A.65}$$

By (A.61) and (A.65), we have

$$||V_{NT,13}(\widehat{f}^{[0]})||=O_p(d_{NT})+o_p(N^{-1/2}). \tag{A.66}$$

Therefore, the result in Lemma 16 follows from (A.59), (A.60), and (A.66) directly. □

Proofs of Theorems 1 and 2. Based on (A.49) in Lemma 14, the result in Lemma 10 holds for $\widehat{f}_t^{[1]}$ with a different bounded sequence. Then the result (A.49) in Lemma 14 holds for $\widehat{g}_j^{[2]}(x_j)$. This process can be continued for any finite number of iterations. By assuming that the algorithm in Section 3.1 converges at the $(i+1)^{\text{th}}$ step for any finite number i, the results in Lemmas 10 and 14 hold for $\widehat{f}_t=\widehat{f}_t^{[i+1]}$ and $\widehat{g}_j=\widehat{g}_j^{[i+1]}(x_j)$. Hence, Theorem 1 for $\widehat{f}_t$ follows from Lemmas 9 and 10, directly, and Theorem 2 for $\widehat{g}_j$ is proved by using Lemma 14. □

8.3 Proofs of Theorem 3

Proof. We prove the consistency of $\widehat{\Lambda}_{Nt}$. Define

$$\widetilde{\Lambda}_{Nt}=(Nh)^{-1}\sum_{i=1}^{N}K\left(\frac{y_{it}-(f_{ut}^0+\sum_{j=1}^{J}g_j^0(X_{ji})f_{jt}^0)}{h}\right)Q_i^0(X_i)Q_i^0(X_i)^\intercal,$$

and

$$\widehat{\Lambda}_{Nt}=(Nh)^{-1}\sum_{i=1}^{N}K\left(\frac{y_{it}-(\widehat{f}_{ut}+\sum_{j=1}^{J}\widehat{g}_j(X_{ji})\widehat{f}_{jt})}{h}\right)\widehat{Q}_i(X_i)\widehat{Q}_i(X_i)^\intercal.$$

We will show $||\widehat{\Lambda}_{Nt}-\widetilde{\Lambda}_{Nt}||=o_p(1)$ and $||\widetilde{\Lambda}_{Nt}-\Lambda_{Nt}^0||=o_p(1)$, respectively. Let $\widehat{d}_{it}(X_i)=\{\widehat{f}_{ut}+\sum_{j=1}^{J}\widehat{g}_j(X_{ji})\widehat{f}_{jt}\}-\{f_{ut}^0+\sum_{j=1}^{J}g_j^0(X_{ji})f_{jt}^0\}$. Then,

$$\widehat{\Lambda}_{Nt}-\widetilde{\Lambda}_{Nt}=D_{Nt,1}+D_{Nt,2},$$

where

$$D_{Nt,1}=(2Nh)^{-1}\sum_{i=1}^{N}\{I(|\varepsilon_{it}|\le h)-I(|\varepsilon_{it}-\widehat{d}_{it}(X_i)|\le h)\}Q_i^0(X_i)Q_i^0(X_i)^\intercal,$$
$$D_{Nt,2}=(2Nh)^{-1}\sum_{i=1}^{N}I(|\varepsilon_{it}-\widehat{d}_{it}(X_i)|\le h)\{\widehat{Q}_i(X_i)\widehat{Q}_i(X_i)^\intercal-Q_i^0(X_i)Q_i^0(X_i)^\intercal\}.$$

Since there exist some constants $0<c_f,c_1<\infty$ such that with probability approaching 1,

$$E\{\widehat{d}_{it}(X_i)\}^2=\int\widehat{d}_{it}^2(x)f_{X_i}(x)dx\le c_f\int\widehat{d}_{it}^2(x)dx\le c_1\phi_{NT}^2+o(N^{-1}),$$

where ϕ_{NT} is given in (4.5), and the last inequality follows from the result in Theorem 2, then there exists some constant $0 < c < \infty$ such that with probability approaching 1,

$$\begin{aligned}E||\widehat{\Lambda}_{Nt} - \widetilde{\Lambda}_{Nt}|| &\le c(2Nh)^{-1}\sum\nolimits_{i=1}^{N} E|\widehat{d}_{it}(X_i)| \times ||Q_i^0(X_i)Q_i^0(X_i)^{\intercal}|| \\ &\le c(2Nh)^{-1}\sum\nolimits_{i=1}^{N} E\{\widehat{d}_{it}(X_i)\}^2 E||Q_i^0(X_i)Q_i^0(X_i)^{\intercal}||^2\}^{1/2} \\ &\le cc_1^{1/2}(2Nh)^{-1}(\sqrt{K_N/(NT)} + K_N^{3/2}N^{-3/4}\sqrt{\log N} + K_N^{-r})\times \\ &\sum\nolimits_{i=1}^{N}\{E||Q_i^0(X_i)Q_i^0(X_i)^{\intercal}||^2\}^{1/2}.\end{aligned}$$

By Condition (C3), we have $\sup_{x_j\in[a,b]}|g_j^0(x_j)| \le C'$ for all j, for any vector $\mathbf{a} \in R^{J+1}$ and $||\mathbf{a}||^2 = 1$, we have

$$\begin{aligned}\mathbf{a}^{\intercal}Q_i^0(X_i)Q_i^0(X_i)^{\intercal}\mathbf{a} &= \{a_0 + \sum\nolimits_{j=1}^{J} g_j^0(X_{ji})a_j\}^2 \le (J+1)\{a_0^2 + g_j^0(X_{ji})^2 a_j^2\} \\ &\le (J+1)\{a_0^2 + (C')^2 a_j^2\} \le C_a\end{aligned}$$

for some constant $0 < C_a < \infty$. Hence, $||Q_i^0(X_i)Q_i^0(X_i)^{\intercal}|| \le C_a$, and thus we have

$$\begin{aligned}E||\widehat{\Lambda}_{Nt} - \widetilde{\Lambda}_{Nt}|| &\le cc_1^{1/2}(2Nh)^{-1}(\phi_{NT} + o(N^{-1/2}))\sum\nolimits_{i=1}^{N} C_a \\ &= 2^{-1}cc_1^{1/2}C_a h^{-1}(\phi_{NT} + o(N^{-1/2})) = o(1)\end{aligned}$$

by the assumption that $h^{-1}\phi_{NT} = o(1)$ and $h^{-1}N^{-1/2} = O(1)$. Hence, we have $||D_{Nt,1}|| = o_p(1)$. Moreover, for any vector $\mathbf{a} \in R^{J+1}$ and $||\mathbf{a}||^2 = 1$, we have with probability approaching 1, there exists a constant $0 < C < \infty$ such that

$$\begin{aligned}|\mathbf{a}^{\intercal}D_{Nt,2}\mathbf{a}| &\le (2Nh)^{-1}\sum\nolimits_{i=1}^{N}|\{a_0 + \sum\nolimits_{j=1}^{J}\widehat{g}_j(X_{ji})a_j\}^2 - \{a_0 + \sum\nolimits_{j=1}^{J} g_j^0(X_{ji})a_j\}^2| \\ &\le C(2Nh)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{J}|\{\widehat{g}_j(X_{ji}) - g_j^0(X_{ji})\}a_j| \\ &\le C(2h)^{-1}\sum\nolimits_{j=1}^{J}\{N^{-1}\sum\nolimits_{i=1}^{N}\{\widehat{g}_j(X_{ji}) - g_j^0(X_{ji})\}^2 a_j^2\}^{1/2} \\ &= O(h^{-1})\{O(\phi_{NT}) + o(N^{-1/2})\} = o(1).\end{aligned}$$

Hence, we have $||D_{Nt,2}|| = o_p(1)$. Therefore, $||\widehat{\Lambda}_{Nt} - \widetilde{\Lambda}_{Nt}|| \le ||D_{Nt,1}|| + ||D_{Nt,2}|| = o_p(1)$. Next, we will show $||\widetilde{\Lambda}_{Nt} - \Lambda_{Nt}^0|| = o_p(1)$. Since

$$\begin{aligned}&|E\left\{(2h)^{-1}I(|\varepsilon_{it}| \le h) - p_i\left(0\,|X_i\right)|X_i, f_t\right\}| \\ &= |(2h)^{-1}h\{p_i\left(h^*\,|X_i, f_t\right) + p_i\left(-h^{**}\,|X_i, f_t\right)\} - p_i\left(0\,|X_i, f_t\right)| \\ &= |2^{-1}[\{p_i\left(h^*\,|X_i, f_t\right) - p_i\left(0\,|X_i, f_t\right)\} + \{p_i\left(-h^{**}\,|X_i, f_t\right) - p_i\left(0\,|X_i, f_t\right)\}]| \le c'h\end{aligned}$$

for some constant $0 < c' < \infty$, where h^* and h^{**} are some values between 0 and h, and the last inequality follows from Condition (C2), then by the above result and Condition (C5),

$$\begin{aligned}||E(\widetilde{\Lambda}_{Nt} - \Lambda_{Nt}^0)|| &= ||N^{-1}\sum\nolimits_{i=1}^{N} E[\{(2h)^{-1}I(|\varepsilon_{it}| \le h) - p_i\left(0\,|X_i, f_t\right)\}Q_i^0(X_i)Q_i^0(X_i)^{\intercal}]|| \\ &\le c'h||N^{-1}\sum\nolimits_{i=1}^{N} EQ_i^0(X_i)Q_i^0(X_i)^{\intercal}|| = O(h) = o(1). \qquad (A.67)\end{aligned}$$

Moreover, by Conditions (C1), we have $E\{I(|\varepsilon_{it}| \leq h)\} \leq 2C^*h$ for some constant $C^* \in (0, \infty)$, and then for any vector $\mathbf{a} \in R^{(J+1)}$ with $||\mathbf{a}|| =1$, by Conditions (C1), (C2) and (C3), we have

$$\begin{aligned}
&\text{var}(\mathbf{a}^\intercal \widetilde{\Lambda}_{Nt}\mathbf{a}) \\
&= (2Nh)^{-2}\text{var}\left(\sum\nolimits_{i=1}^{N} I(|\varepsilon_{it}| \leq h)\{a_0 + \sum\nolimits_{j=1}^{J} g_j^0(X_{ji})a_j\}^2\right) \\
&\leq (2Nh)^{-2}\sum\nolimits_{i,i'} 2\{\phi(|i-i'|)\}^{1/2}\times \\
&\left(E\left[I\left(|\varepsilon_{it}| \leq h\right)\{a_0 + \sum\nolimits_{j=1}^{J} g_j^0(X_{ji})a_j\}^4\right]\right)^{1/2}\left(E\left[I(|\varepsilon_{i't'}| \leq h)\{a_0 + \sum\nolimits_{j=1}^{J} g_j^0(X_{ji'})a_j\}^4\right]\right)^{1/2} \\
&\leq (J+1)^2\{a_0^2 + C'^2 a_j^2\}(2Nh)^{-2}(2C^*h)^2\sum\nolimits_{i,i'} 2\{\phi(|i-i'|)\}^{1/2} \\
&\leq (J+1)^2\{a_0^2 + C'^2 a_j^2\}N^{-2}2C^{*2}K_1\sum\nolimits_{i,i'} e^{-(\lambda_1/2)(|i-i'|)} \\
&\leq (J+1)^2\{a_0^2 + C'^2 a_j^2\}2C^{*2}K_1N^{-1}\{1-e^{-(\lambda_1/2)}\} = O(N^{-1}) = o(1). \qquad \text{(A.68)}
\end{aligned}$$

By (A.67) and (A.68), we have $||\widetilde{\Lambda}_{Nt} - \Lambda_{Nt}^0|| = o_p(1)$. Hence, $||\widehat{\Lambda}_{Nt} - \Lambda_{Nt}^0|| \leq ||\widehat{\Lambda}_{Nt} - \widetilde{\Lambda}_{Nt}|| + ||\widetilde{\Lambda}_{Nt} - \Lambda_{Nt}^0|| = o_p(1)$. □

8.4 Proofs of Theorem 4

Proof. Let $S_{[rN]t} = \sum_{i=1}^{[rN]} Q_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))$, where $[a]$ denotes the largest integer no greater than a. Let $M = bN$. Define $\Lambda_{Nt}(r) = N^{-1}\sum_{i=1}^{[rN]} p_i\left(0\,|X_i, f_t\right) Q_i^0(X_i)Q_i^0(X_i)^\intercal$, $\digamma_{Nt}(r) = N^{-1/2}S_{[rN]t}$, and

$$D_{bN}(r) = N^2\left(K^*\left(\frac{[rN]+1}{bN}\right) - K^*\left(\frac{[rN]}{bN}\right)\right) - \left(K^*\left(\frac{[rN]}{bN}\right) - K^*\left(\frac{[rN]-1}{bN}\right)\right).$$

Denote $K_{ij}^* = K^*(\frac{i-j}{bN})$, and $\widehat{w}_{Nt} = \frac{\tau(1-\tau)}{N}\sum_{i=1}^{N}\widehat{Q}_i(X_i)\widehat{Q}_i(X_i)^\intercal - N^{-1}\sum_{i=1}^{N}\widehat{v}_{it}\widehat{v}_{it}^\intercal$. Then

$$\begin{aligned}
\widehat{\Omega}_{Nt,N} &= N^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\widehat{v}_{it}K_{ij}^*\widehat{v}_{jt}^\intercal + \widehat{w}_{Nt} \\
&= N^{-1}\sum\nolimits_{i=1}^{N}(\widehat{v}_{it}\sum\nolimits_{j=1}^{N}K_{ij}^*\widehat{v}_{jt}^\intercal) + \widehat{w}_{Nt}.
\end{aligned}$$

Define $\widehat{S}_{nt} = \sum_{i=1}^{n}\widehat{v}_{it}$. By the assumptions in Theorem 1, $\phi_{NT}N^{1/2} = o(1)$ and by the results in Lemmas 9-16, we have

$$\widehat{f}_t - f_t^0 = \Lambda_{Nt}^{-1}\{N^{-1}\sum\nolimits_{i=1}^{N} Q_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\} + o_p(N^{-1/2}), \qquad \text{(A.69)}$$

$$\sup_{x_j \in \mathcal{X}_j} |\widehat{g}_j(x_j) - g_j^0(x_j)| = O_p(\phi_{NT}) + o_p(N^{-1/2}) = o_p(N^{-1/2}). \qquad \text{(A.70)}$$

Let $r \in (0,1]$. Let $\widetilde{S}_{[rN]t} = \sum_{i=1}^{[rN]} Q_i^0(X_i)(\tau - I(\widehat{\varepsilon}_{it}^0 < 0))$, where $\widehat{\varepsilon}_{it}^0 = y_{it} - \{\widehat{f}_{ut} + \sum_{j=1}^{J} g_j^0(X_{ji})\widehat{f}_{jt}\}$. By Lemma 13, we have

$$||N^{-1/2}\widehat{S}_{[rN]t} - N^{-1/2}\widetilde{S}_{[rN]t}|| = o_p(1). \qquad \text{(A.71)}$$

For any given $f_t \in R^{J+1}$, define $S_{[rN]t}(f_t) = \sum_{i=1}^{[rN]} Q_i^0(X_i)(\tau - I(\varepsilon_{it}(f_t) < 0))$, where $\varepsilon_{it}(f_t) = y_{it} - \{f_{ut} + \sum_{j=1}^{J} g_j^0(X_{ji})f_{jt}\}$. Following similar arguments to the proof in Lemma 16, we have

$$\sup_{||f_t - f_t^0|| \leq C(d_{NT}+N^{-1/2})} ||N^{-1/2}[S_{[rN]t}(f_t) - S_{[rN]t}(f_t^0) - E[\{S_{[rN]t}(f_t) - S_{[rN]t}(f_t^0)\}|\mathbf{X}, \mathbf{F}]]|| = o_p(1).$$

Moreover,

$$
\begin{aligned}
&N^{-1/2}E[\{S_{[rN]t}(f_t) - S_{[rN]t}(f_t^0)\}|\mathbf{X}, \mathbf{F}] \\
&= \sum\nolimits_{i=1}^{[rN]} Q_i^0(X_i) E[(I(\varepsilon_{it}(f_t^0) < 0) - I(\varepsilon_{it}(f_t) < 0))|X_i, f_t],
\end{aligned}
\tag{A.72}
$$

and thus by Taylor's expansion, we have

$$
\begin{aligned}
&||N^{-1/2}E[\{S_{[rN]t}(f_t) - S_{[rN]t}(f_t^0)\}|\mathbf{X}, \mathbf{F}] \\
&-N^{-1/2}\sum\nolimits_{i=1}^{[rN]} p_i\,(0\,|X_i, f_t\,)\, Q_i^0(X_i) Q_i^0(X_i)^{\intercal}(f_t^0 - f_t)|| = o_p(1).
\end{aligned}
\tag{A.73}
$$

Hence, by (A.71), (A.72) and (A.73), we have

$$
\begin{aligned}
N^{-1/2}\widehat{S}_{[rN]t} &= N^{-1/2}\sum\nolimits_{i=1}^{[rN]} Q_i^0(X_i)(\tau - I(\varepsilon_{it} < 0)) \\
&\quad - N^{-1/2}\sum\nolimits_{i=1}^{[rN]} p_i\,(0\,|X_i, f_t\,)\, Q_i^0(X_i) Q_i^0(X_i)^{\intercal}(\widehat{f}_t - f_t^0) + o_p(1).
\end{aligned}
$$

This result, together with (A.69), implies

$$
N^{-1/2}\widehat{S}_{[rN]t} = \digamma_{Nt}(r) - \Lambda_{Nt}(r)\{\Lambda_{Nt}(1)\}^{-1}\digamma_{Nt}(1) + o_p(1). \tag{A.74}
$$

Thus, $N^{-1/2}\widehat{S}_{Nt} = o_p(1)$. By following the argument above again, we have $||N^{-1/2}\sum_{j=1}^{N}\widehat{v}_{jt}K_{jN}^* - N^{-1/2}\sum_{j=1}^{N} v_{jt}K_{jN}^*|| = O_p(1)$. Also $||N^{-1/2}\sum_{j=1}^{N} v_{jt}K_{jN}^*|| = O_p(1)$ by the weak law of large numbers. Hence, $||N^{-1/2}\sum_{j=1}^{N}\widehat{v}_{jt}K_{jN}^*|| = O_p(1)$. Therefore

$$
N^{-1}\sum\nolimits_{j=1}^{N}\widehat{v}_{jt}K_{jN}^*\widehat{S}_N^{\intercal} = O_p(1)o_p(1) = o_p(1).
$$

By (A.69) and (A.70), $\widehat{w}_{Nt} = o_p(1)$. By this result and also applying the identity that $\sum_{l=1}^{N} a_l b_l = (\sum_{l=1}^{N-1}(a_l - a_{l+1})\sum_{j=1}^{l} b_j) + a_N\sum_{l=1}^{N} b_l$ to $\sum_{j=1}^{N} K_{ij}^*\widehat{v}_j^{\intercal}$ and then again to the sum over i, we obtain

$$
\begin{aligned}
\widehat{\Omega}_{Nt,M=bN} &= N^{-1}\sum\nolimits_{i=1}^{N-1} N^{-1}\sum\nolimits_{j=1}^{N-1} N^2((K_{ij}^* - K_{i,j+1}^*) - (K_{i+1,j}^* - K_{i+1,j+1}^*))N^{-1/2}\widehat{S}_{it}N^{-1/2}\widehat{S}_{jt}^{\intercal} \\
&\quad + N^{-1}\sum\nolimits_{j=1}^{N}\widehat{v}_{jt}K_{jN}^*\widehat{S}_{Nt}^{\intercal} + o_p(1),
\end{aligned}
$$

and thus

$$
\begin{aligned}
\widehat{\Omega}_{Nt,M=bN} &= \sum\nolimits_{i=1}^{N-1}\sum\nolimits_{j=1}^{N-1}((K_{ij}^* - K_{i,j+1}^*) - (K_{i+1,j}^* - K_{i+1,j+1}^*))\frac{\widehat{S}_{it}}{\sqrt{N}}\frac{\widehat{S}_{jt}^{\intercal}}{\sqrt{N}} \\
&+ o_p(1).
\end{aligned}
\tag{A.75}
$$

Moreover,

$$
N^2((K_{ij}^* - K_{i,j+1}^*) - (K_{i+1,j}^* - K_{i+1,j+1}^*)) = -D_{bN}\{(i-j)/N\}. \tag{A.76}
$$

Also $\lim_{N\to\infty} D_{bN}(r) = \frac{1}{b^2}K^{*\prime\prime}(\frac{r}{b})$, $||\Lambda_{Nt}(r) - r\Lambda_t^0|| = o_p(1)$, where $\Lambda_t^0 = \lim_{N\to\infty}\Lambda_{Nt}^0$ and $\digamma_{Nt}(r) \xrightarrow{\mathcal{D}} W_{J+1}(r)\Upsilon^{\intercal}$. Thus,

$$
(\Lambda_{Nt}(r), \digamma_{Nt}(r)^{\intercal}, D_{bN}(r)) \xrightarrow{\mathcal{D}} \left(r\Lambda_t^0, \Upsilon W_{J+1}(r)^{\intercal}, \frac{1}{b^2}K^{*\prime\prime}\left(\frac{r}{b}\right)\right). \tag{A.77}
$$

Hence, by (A.74), (A.75), and (A.76), it follows that

$$\widehat{\Omega}_{Nt,M=bN} = \int_0^1 \int_0^1 -D_{bN}(r-s)[\digamma_{Nt}(r) - \Lambda_{Nt}(r)\{\Lambda_{Nt}(1)\}^{-1}\digamma_{Nt}(1)] \times [\digamma_{Nt}(s) - \Lambda_{Nt}(s)\{\Lambda_{Nt}(1)\}^{-1}\digamma_{Nt}(1)]^{\intercal} drds + o_p(1). \tag{A.78}$$

By the continuous mapping theorem,

$$\widehat{\Omega}_{N,M=bN} \xrightarrow{\mathcal{D}} \Upsilon \int_0^1 \int_0^1 -\frac{1}{b^2}K^{*\prime\prime}(\frac{r-s}{b})\{W_{J+1}(r) - rW_{J+1}(1)\}\{W_{J+1}(s) - sW_{J+1}(1)\}^{\intercal} drds\Upsilon^{\intercal}.$$

Then the proof is completed. □

8.5 Proofs of Theorems 5 and 6

Proof. By (A.69), $\widehat{f}_t - f_t^0 = N^{-1/2}\Lambda_{Nt}(1)^{-1}\digamma_{Nt}(1) + o_p(N^{-1/2})$. Then under H_0, we have

$$N^{1/2}(R\widehat{f}_t - r) = R\Lambda_{Nt}(1)^{-1}\digamma_{Nt}(1) + o_p(1). \tag{A.79}$$

It directly follows from (A.77), (A.78) and (A.79) that

$$\begin{aligned} F_{Nt,b} &\xrightarrow{\mathcal{D}} \{R\Lambda_t^{0-1}\Upsilon W_{J+1}(1)\}^{\intercal}\{R\tau(1-\tau)\Lambda_t^{0-1} \\ &\times (\Upsilon \int_0^1 \int_0^1 -\frac{1}{b^2}K^{*\prime\prime}(\frac{r-s}{b})B_{J+1}(r)B_{J+1}(s)^{\intercal} drds\Upsilon^{\intercal})\Lambda_t^{0-1}R^{\intercal}\}^{-1} \\ &\times R\Lambda_t^{0-1}\Upsilon W_{J+1}(1)/q. \end{aligned}$$

Since $R\Lambda_t^{0-1}\Upsilon W_{J+1}(1)$ is a $q \times 1$ vector of normal random variables with mean zero and variance $R\Lambda_t^{0-1}\Upsilon\Upsilon^{\intercal}\Lambda_t^{0-1}R^{\intercal}$, $R\Lambda_t^{0-1}\Upsilon W_{J+1}(1)$ can be written as $\Upsilon_t^* W_q(1)$, where $\Upsilon_t^*\Upsilon_t^{*\intercal} = R\Lambda_t^{0-1}\Upsilon\Upsilon^{\intercal}\Lambda_t^{0-1}R^{\intercal}$. Then replacing $R\Lambda_t^{0-1}\Upsilon W_{J+1}(1)$ by $\Upsilon_t^* W_q(1)$ and canceling Υ_t^* in the above equation, we have the result in Theorem 5. Moreover, under the alternative that H_1: $Rf_t^0 = r + cN^{-1/2}$, we have

$$\begin{aligned} N^{1/2}(R\widehat{f}_t - r) &= N^{1/2}(Rf_t^0 - r) + R\Lambda_{Nt}(1)^{-1}\digamma_{Nt}(1) + o_p(1) \\ &= c + R\Lambda_{Nt}(1)^{-1}\digamma_{Nt}(1) + o_p(1). \end{aligned}$$

Thus by (A.77), we have

$$\begin{aligned} F_{Nt,b} &\xrightarrow{\mathcal{D}} \{c + R\Lambda_t^{0-1}\Upsilon W_{J+1}(1)\}^{\intercal}\{R\tau(1-\tau)\Lambda_t^{0-1} \\ &\times (\Upsilon \int_0^1 \int_0^1 -\frac{1}{b^2}K^{*\prime\prime}(\frac{r-s}{b})B_{J+1}(r)B_{J+1}(s)^{\intercal} drds\Upsilon^{\intercal})\Lambda_t^{0-1}R^{\intercal}\}^{-1} \\ &\times \{c + R\Lambda_t^{0-1}\Upsilon W_{J+1}(1)\}/q. \end{aligned}$$

Also $c + R\Lambda_t^{0-1}\Upsilon W_{J+1}(1) \equiv c + \Upsilon_t^* W_q(1) = \Upsilon_t^*(\Upsilon_t^{*-1}c + W_q(1))$. Then the result in Theorem 6 follows from the above results. The proof is completed.

□

References

[1] Angrist, J., and J. Pischke (2009): "Mostly Harmless Econometrics: An Empiricist's Companion," 1st Edition, New Jersey: Princeton University Press.

[2] Babu, G. J. (1989): "Strong Representations for LAD Estimators in Linear Models," *Probability Theory and Related Fields*, 83, 547-558.

[3] Bai, J. and Ng, S. (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70,191221.

[4] Banz, R.W. (1981): "The Relationship between Return and Market Value of Common Stocks," *Journal of Financial Economics*, 9, 3-18.

[5] Bassett., G.W., R. Koenker and G. Kordas (2004): "Pessimistic Portfolio Allocation and Choquet Expected Utility," *Journal of Financial Econometrics*, 2, 477-492.

[6] Basu, S. (1977): "The Investment Performance of Common Stocks in Relation to Their Price to Earnings Ratio: a Test of the Efficient Markets Hypothesis," *Journal of Finance*, 32, 663-682.

[7] Belloni, A., V. Chernozhukov, and I. Fernandez-Val (2016): "Conditional Quantile Processes based on Series or Many Regressors," Available at arXiv:1105.6154v1.

[8] Boneva, L., O. Linton, and M. Vogt (2015): "The Effect of Fragmentation in Trading on Market Quality in the UK Equity Market," *Journal of Applied Econometrics*, Available at http://dx.doi.org/10.1002/jae.2438

[9] Bosq, D. (1998): "Nonparametric Statistics for Stochastic Processes," New York: Springer-Verlag.

[10] Brown, S. J. (1989): "The Number of Factors in Security Returns," *Journal of Finance*, 44, 1247-1262.

[11] Bryzgalova, S. (2015): "Spurious Factors in Linear Asset Pricing Models. Working Paper.

[12] Chen, X. (2011): "Penalized Sieve Estimation and Inference of Semi–Nonparametric Dynamic Models: A Selective Review," Cowles Foundation Discussion Paper No. 1804.

[13] Chen, X., and T. Christensen (2015): "Optimal Uniform Convergence Rates and Asymptotic Normality for Series Estimators under Weak Dependence and Weak Conditions," *Journal of Econometrics*, 188, 447–465.

[14] Chen, X., and Z. Liao (2012): "Asymptotic Properties of Penalized M Estimators with Time Series Observations," In *Causality, Prediction and Specification Analysis: Essays in Honour of Halbert White* (eds. X. Chen and N. Swanson). Springer–Verlag, Berlin.

[15] Chen, X., and D. Pouzo (2015): "Sieve Quasi–Likelihood Ratio Inference on Semi–Nonparametric Conditional Moment Models," *Econometrica*, 83, 1013–1079.

[16] Conley, T. G. (1999): "GMM Estimation with Cross–Sectional Dependence," *Journal of Econometrics*, 92, 1-45.

[17] Connor, G., and R.A. Korajczyk (1993): "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263-1288.

[18] Connor, G., and O. Linton (2007): "Semiparametric Estimation of a Characteristic–Based Factor Model of Stock Returns," *Journal of Empirical Finance*, 14, 694-717.

[19] Connor, G., M. Hagmann, and O. Linton (2012): "Efficient Semiparametric Estimation of the Fama–French Model and Extensions," *Econometrica*, 80, 713-754.

[20] Daniel, K., and S. Titman (1997): "Evidence on the Characteristics of Cross–sectional Variation in Stock Returns," *Journal of Finance*, 52, 1-34.

[21] Davis, J. (1994): "The Cross–Section of Realized Stock Returns: the pre-Compustat Evidence," *Journal of Finance*, 49, 1579-1593.

[22] de Boor, C. (2001): A Practical Guide to Splines," Applied Mathematical Sciences 27, New York: Springer.

[23] Demko, S. (1986): "Spectral Bounds for $|a^{-1}|_\infty$," *Journal of Approximation Theory*, 48, 207-212.

[24] Dong, C., J. Gao, and B. Peng (2015): "Semiparametric Single–Index Panel Data Models with Cross–Sectional Dependence," *Journal of Econometrics*, 188, 301–312

[25] Fama, E.F., and K.R. French (1992): "The Cross–Section of Expected Stock Returns," *Journal of Finance*, 47, 427-465.

[26] Fama, E.F., and K.R. French (1993): "Common Risk Factors in the Returns to Stocks and Bonds," *Journal of Financial Economics*, 33, 3-56.

[27] Fama, E.F., and K.R. French (1995): "Size and Book to Market Factors in Earnings and Returns," *Journal of Finance*, 50, 131-156.

[28] Fama, E.F., and K.R. French (1996): "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55-84.

[29] Fama, E.F., and K.R. French (1998): "Value versus Growth: the International Evidence," *Journal of Finance*, 53, 1975-2000.

[30] Fan, J., Y. Liao, and M. Micheva (2013): "Large Covariance Estimation by Thresholding Principal Orthogonal Complements (with Discussion)," *Journal of the Royal Statistical Society Series B*, 75, 603–680.

[31] Fan, J., Y. Liao, and W. Wang (2016): "Projected Principal Component Analysis in Factor Models," *The Annals of Statistics*, 44, 219-254.

[32] Gao, J., Z. Lu, and D. Tjøstheim (2006): "Estimation in Semiparametric Spatial Regression," *The Annals of Statistics*, 34, 1395-1435.

[33] Haugen, R. (1995): "The New Finance: the case against Efficient Markets," New Jersey: Prentice-Hall, Englewood Cliffs.

[34] Hodrick, R., D. Ng, and P. Sengmueller (1999): "An International Dynamic Asset Pricing Model," *International Taxation and Public Finance*, 6, 597-620.

[35] Horowitz, J. L., and Lee, S. (2005): "Nonparametric Estimation of an Additive Quantile Regression Model," *Journal of the American Statistical Association*, 100, 1238-1249.

[36] Horowitz, J. L., and Mammen, E. (2011): "Oracle–Efficient Nonparametric Estimation of an Additive Model with an Unknown Link Function," *Econometric Theory*, 27, 582–608.

[37] Kiefer, N. M., and T. J. Vogelsang (2002): "Heteroskedasticity–Autocorrelation Robust Standard Errors using the Bartlett Kernel without Truncation," *Econometrica,* 70, 2093-2095.

[38] Kiefer, N. M., and T. J. Vogelsang (2005): "A New Asymptotic Theory for Heteroskedasticity–Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130-1164.

[39] Knight, K. (1998): "Limiting Distribution for L_1 Regression Estimators under General Conditions," *The Annals of Statistics*, 26, 755-770.

[40] Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33-50.

[41] Lakonishok, J., A. Shleifer, and R.W. Vishny (1994): "Contrarian Investment, Extrapolation and Risk," *Journal of Finance*, 49, 1541-1578.

[42] Lee, J., and Robinson, P. M. (2016): "Series Estimation under Cross–Sectional Dependence," *Journal of Econometrics*, 190, 1–17.

[43] Lewellen, J. (1999): "The Time–Series Relations among Expected Return, Risk, and Book to Market Value," *Journal of Financial Economics*, 54, 5-44.

[44] Ma, S., and X. He (2016): "Inference for Single–Index Quantile Regression Models with Profile Optimization," *The Annals of Statistics*, forthcoming.

[45] Ma, S., Q. Song, and L. Wang (2013): "Simultaneous Variable Selection and Estimation in Semiparametric Modelling of Longitudinal/clustered Data," *Bernoulli*, 19, 252-274.

[46] Ma, S., and L. Yang (2011): "Spline–Backfitted Kernel Smoothing of Partially Linear Additive Model," *Journal of Statistical Planning and Inference*, 14, 204-219.

[47] MacKinlay, A.C. (1995): "Multifactor Models do not Explain Deviations from the CAPM," *Journal of Finance*, 38, 3-23.

[48] Merlevède, F., M. Peligrad, and E. Rio (2009): "Bernstein Inequality and Moderate Deviations under Strong Mixing Conditions," *IMS Collections, High Dimensional Probability V: The Luminy Volume*, 5, 273-292.

[49] Pesaran, M.H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967-1012.

[50] Powell, J. L. (1991): "Estimation of Monotonic Regression Models under Quantile Restrictions," In W. Barnett, J. Powell, G. Tauchen (Eds.), Nonparametric and Semiparametric Models in Econometrics, Cambridge: Cambridge University Press.

[51] Robinson, P.M., and S. Thawornkaiwong (2012): "Statistical Inference on Regression with Spatial Dependence," *Journal of Econometrics*, 167, 521-542.

[52] Rosenberg, B. (1974): Extra–Market Components of Covariance Among Security Prices," *Journal of Financial and Quantitative Analysis*, 9, 263-274.

[53] Rosenberg, B., K. Reid, and R. Lanstein (1985): "Persuasive Evidence of Market Inefficiency," *Journal of Portfolio Management*, 11, 9-17.