UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1933

ESTIMATION AND INFERENCE IN SEMIPARAMETRIC QUANTILE FACTOR MODELS

Shujie Ma   Oliver Linton   Jiti Gao

25 March 2019

We consider a semiparametric quantile factor panel model that allows observed stock-specific characteristics to affect stock returns in a nonlinear time-varying way, extending Connor, Hagmann, and Linton (2012) to the quantile restriction case. We propose a sieve-based estimation methodology that is easy to implement. 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 daily stock return data where we find significant evidence of nonlinearity in many of the characteristic exposure curves.

Estimation and Inference in Semiparametric Quantile Factor Models

Shujie Ma
Department of Statistics, University of California at Riverside
Oliver Linton
Department of Economics, University of Cambridge
Jiti Gao
Department of Econometrics and Business Statistics, Monash University

Abstract

We consider a semiparametric quantile factor panel model that allows observed stock-specific characteristics to affect stock returns in a nonlinear time-varying way, extending Connor, Hagmann, and Linton (2012) to the quantile restriction case. We propose a sieve-based estimation methodology that is easy to implement. 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 daily stock return data where we find significant evidence of nonlinearity in many of the characteristic exposure curves.

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

Short title: Semiparametric Quantile Factor Models

Shujie Ma is Associate Professor, Department of Statistics, University of California at Riverside, Riverside, CA 92521 (Email: shujie.ma@ucr.edu). Oliver Linton is Professor, Department of Economics, University of Cambridge, Cambridge CB3 9DD, UK (Email: obl20@cam.ac.uk). Jiti Gao is Professor, Department of Econometrics and Business Statistics, Monash University, Victoria 3145, Australia (Email: jiti.gao@monash.edu). The research of Ma is supported in part by the U.S. NSF grant DMS 1712558. The research of Gao is supported in part by the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421.

1 Introduction

Factor models are widely used to capture the co-movement of a large number of time series and to model covariance matrices. They provide useful dimensionality reduction in many applications from climate modelling to finance. Perhaps the current state of the art for factor modelling is Fan, Liao, and Micheva (2013), which allowed the idiosyncratic covariance matrix to be non-diagonal but sparse, and used thresholding techniques (Cai and Liu, 2011) to impose sparsity and thereby obtain a better estimator of the covariance matrix and its inverse in this big-data setting. The usual approach ignores covariate information that can sometimes be informative. Connor, Hagmann and Linton (2012) developed a semiparametric factor regression methodology that introduces covariate information into the factor loading parameters. This model is well motivated in finance applications where it can be understood as a properly formulated version of the popular Fama-French (1992) approach to modelling returns with observable characteristics. The model also makes sense in other contexts where covariate information is available. Their application was to monthly stock returns, which is where the finance literature was focussed. Moreover, Fan, Liao and Wang (2016) proposed a Projected-PCA approach which employs principal component analysis to the projected data matrix onto a linear space spanned by covariates. It is worth noting that most existing works in the literature of factor models require at least four moments to establish their theoretical properties. See, for instance, Bai and Ng (2002), Bai and Li (2012), Lam and Yao (2012), Connor, Hagmann and Linton (2012), Fan, Liao, and Micheva (2013), Fan, Liao and Wang (2016), Li et al. (2017), among others. This may not be a binding restriction for monthly stock returns, but for daily stock returns this is a bit strong.

Quantile methods are widely used in statistics. They have the advantage of being robust to large observations. They can also provide more information about the conditional distribution away from the centre, which is relevant in many applications. In this paper, we propose estimation and inferential methodology for the quantile version of the Connor, Hagmann and Linton (2012) model. 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 use these to estimate the factor return vector at each time period. We then update the loading functions and factor returns 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 simultaneous estimation of these parameters without penalization would be challenging. Penalization of the factor returns here is not 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 average values 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 to be not 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. Conley (1999) studied that under a lattice structure or some observable or estimable distance function that determines the ordering, 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 achieve this, namely, we construct a test statistic based on an inconsistent fixed-b 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 only requires that the time series average value 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 and on-line supplemental materials.

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 $a_n^{-1} b_n = o(1)$, and let $a_n \ll b_n$ denote $a_n b_n^{-1} = o(1)$. For any vector $\mathbf{a} = (a_1, \ldots, a_n)^\intercal \in \mathbb{R}^n$, denote $||\mathbf{a}|| = \left(\sum_{i=1}^n a_i^2\right)^{1/2}$. For any matrix $\mathbf{A}_{m\times n}$, denote its L_2 norm as $\|\mathbf{A}\| = \max_{\boldsymbol{\zeta}\in\mathbb{R}^n, \boldsymbol{\zeta}\neq \mathbf{0}} \|\mathbf{A}\boldsymbol{\zeta}\| \, \|\boldsymbol{\zeta}\|^{-1}$. 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

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

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where y_{it} is the excess return to security i at time t; $f_{ut,\tau}$ and $f_{jt,\tau}$ are factor returns, which are unobservable and treated as fixed-effects parameters to be estimated; $g_{j,\tau}(X_{ji})$ 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 $\mathcal{X}_{ji}$. Let $X_i = (X_{1i}, \ldots, X_{Ji})^\intercal$. Model (2.1) can be written as

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

where the error terms ε_{it} are the asset-specific or idiosyncratic returns and they satisfy that the conditional τ^{th} quantile of ε_{it} given X_i is zero. Note that the factors $f_{ut,\tau}$ and $f_{jt,\tau}$ and the factor betas $g_{j,\tau}(\cdot)$ are τ specific, so is the error term ε_{it}. For notational simplicity, we suppress the τ subscripts such that $f_{ut,\tau} = f_{ut}$, $f_{jt,\tau} = f_{jt}$ and $g_{j,\tau}(\cdot) = g_j$. Let $f_t = (f_{ut}, f_{1t}, \ldots, f_{Jt})^\intercal$. For model identifiability, we assume that:

Assumption A0. $E\{g_j(X_{ji})\} = 0$ and $E\{g_j(X_{ji})\}^2 = 1$ for all $j = 1, \ldots, J$. Furthermore, $\liminf_{T\to\infty} \left|\sum_{t=1}^{T} f_{jt}/T\right| > 0$ for each j.

The case where $\tau = 1/2$ corresponds to the conditional median, and is comparable to the conditional mean model used in 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

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

where $h_{jt,\tau}(\cdot)$ are unknown functions and are τ specific. Again for simplicity, we suppress the τ subscripts by writing $h_{ut,\tau} = h_{ut}$ and $h_{jt,\tau}(\cdot) = h_{jt}(\cdot)$. Without loss of generality, we assume $E\{h_{jt}X_{ji})\} = 0$ for $t = 1, \ldots, T$ (Horowitz and Lee, 2005). Under the factor structure (2.1), we have for all j

$$E\{\frac{1}{T}\sum_{t=1}^{T} h_{jt}(X_{ji})\}^2 = E\{g_j(X_{ji})\}^2 \times (\frac{1}{T}\sum_{t=1}^{T} f_{jt})^2 = (\frac{1}{T}\sum_{t=1}^{T} f_{jt})^2. \tag{2.4}$$

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

$$g_j(x_j) = \frac{\frac{1}{T}\sum_{t=1}^{T} h_{jt}(x_j)}{\sqrt{E\{\frac{1}{T}\sum_{t=1}^{T} h_{jt}(X_{ji})\}^2}}. \tag{2.5}$$

We will use this as the basis for the proposal of the estimation method in Section 3 below.

3 Estimation

3.1 Factor returns and characteristic-beta functions

We propose an alternate optimization algorithm to estimate the factor returns and the characteristic-beta functions. The algorithm makes use of the structure in (2.2) so that it circumvents the "curse of dimensionality" (Bellman, 1961) while retaining flexibility of the nonparametric regression. The right hand side of (2.1) is biconvex in unknown quantities, so it seems difficult to avoid such an algorithmic approach.

To estimate $g_j(\cdot)$, we approximate them by splines. We adopt the centered and standardized B-spline basis functions of order m introduced in Xue and Yang (2006): $B_j(x_j) = \{B_{j,1}(x_j), \ldots, B_{j,K_N}(x_j)\}^\intercal$, satisfying $\mathrm{var}\{B_{jk}(X_j)\} \asymp 1$, where $K_N = L_N + m$, L_N is the number of interior knots satisfying $L_N \to \infty$ as $N \to \infty$. 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. Denote $f_t = \{f_{ut}, (f_{jt}, 1 \leq j \leq J)^\intercal\}^\intercal$. 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 alternate optimization algorithm is described as follows:

1. Find the initial estimates $\widehat{g}_j^{[0]}(\cdot)$.
2. For given $\widehat{g}_j^{[i]}(x_j)$, we obtain for $t = 1, \ldots, T$

$$\widehat{f}_t^{[i+1]} = \arg\min_{f_t \in \mathbb{R}^{J+1}} \sum_{i=1}^{N} \rho_\tau\left(y_{it} - f_{ut} - \sum_{j=1}^{J} \widehat{g}_j^{[i]}(X_{ji}) f_{jt}\right). \tag{3.1}$$

3. For given $\widehat{f}^{[i+1]}$, we obtain

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

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

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

We repeat steps 2 and 3, and let the final estimates be $\widehat{f}_t = \widehat{f}_t^{[i+1]}$ and $\widehat{g}_j(x_j) = \widehat{g}_j^{[i+1]}(x_j)$ if the algorithm is stopped at the $(i+1)^{\text{th}}$ step. We will show in Section 4 that for any finite number i, the estimates $\widehat{f}_t = \widehat{f}_t^{[i+1]}$ and $\widehat{g}_j(x_j) = \widehat{g}_j^{[i+1]}(x_j)$ have desirable asymptotic properties. In our numerical analysis, we let the algorithm stop at the $(i+1)^{\text{th}}$ step according to the two rules: (1) $||\widehat{f}^{[i+1]} - \widehat{f}^{[i]}|| + ||\widehat{\boldsymbol{\lambda}}^{[i+1]} - \widehat{\boldsymbol{\lambda}}^{[i]}|| < \epsilon$ for a small positive value ϵ (we let $\epsilon = 10^{-3}$); (2) setting a finite number i such as $i = 0, 1$ or 2. Our experience in numerical analysis suggests that if the first stopping criterion is used, the algorithm stops after a finite number of iterations by using the consistent initial values proposed in Section 3.2. The estimation is the problem of minimizing a biconvex function such that the objective function is convex in one set of parameters for fixed the other set of parameters. We refer to Gorski et al. (2007) for the convergence property of this alternate optimization algorithm.

3.2 Initial estimators

The proposed iterative algorithm given in Section 3.1 starts from the initial estimates $\widehat{g}_j^{[0]}(\cdot)$ which are obtained by the following way. 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

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

with respect to $(h_{ut}, \boldsymbol{\theta}_t^\intercal)^\intercal \in \mathbb{R}^{JK_N+1}$. As a result, the estimator of $h_{jt}(x_j)$ is $\widetilde{h}_{jt}(x_j) = B_j(x_j)^\intercal \widetilde{\boldsymbol{\theta}}_{jt}$. We then obtain the initial estimators of $g_j(x_j)$

$$\widehat{g}_j^{[0]}(x_j) = \frac{T^{-1}\sum_{t=1}^T \widetilde{h}_{jt}(x_j)}{\sqrt{N^{-1}\sum_{i=1}^N \left(\frac{1}{T}\sum_{t=1}^T \widetilde{h}_{jt}(X_{ji})\right)^2}}. \tag{3.2}$$

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) and Lee and Robinson (2016). Our estimation procedure does not need to know the ordering of the data. However, to develop a robust inference procedure that accounts for heteroscedasticity and cross-sectional correlation (HAC), we need to order the data across i. The ordering assumption can be relaxed by allowing s_N indices to be mis-assigned, where s_N satisfies $s_N = o(N^{1/2})$. We refer to Remark 4 in Section 5 for a detailed discussion. As discussed in Lee and Robinson (2016), in some economic applications, data may be ordered according to some explanatory variables. Such considerations are pursued in our real data analysis with detailed discussions given in Section 7. For notational simplicity, we denote the indices as $\{i, 1 \leq i \leq N\}$ after the ordering.

Let $g_j^0(\cdot)$ for $j = 1, \ldots, J$ and $f_t^0 = (f_{ut}^0, f_{1t}^0, \ldots, f_{Jt}^0)^\intercal$ be the true factor betas and factor returns in model (2.2). For model identifiability, assume $E\{g_j^0(X_{ji})\} = 0$ and $E\{g_j^0(X_{ji})\}^2 = 1$. Let $\mathbb{N}$ denote the collection of all positive integers. We use a ϕ-mixing coefficient to specify the dependence structure. Let $\{W_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$, where $W_{it} = (X_i^\intercal, \varepsilon_{it})^\intercal$ 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

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

where

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

Without loss of generality, we assume that $\mathcal{X}_{ji} = [a, b]$. Denote $h_t^0(x) = \{h_{jt}^0(x_j), 1 \leq j \leq J\}^\intercal$, where $h_{jt}^0(\cdot)$ are the true unknown functions in (2.3) and $x = (x_1, \ldots, x_J)^\intercal$. 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 $\phi(k) \leq K_1 e^{-\lambda_1 k}$ for $K_1, \lambda_1 > 0$.

(C2) The conditional density $p_{it}(\varepsilon\,|x_i)$ of ε_{it} given x_i satisfies the Lipschitz condition of order 1 and $\inf_{1\le i\le N,1\le t\le T} p_{it}(0\,|x_i) > 0$. 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 $m \ge r$.

(C4) There exist some constants $0 < c_h \le C_h < \infty$ such that $c_h \le \left(\frac{1}{T}\sum_{t=1}^T f_{jt}^0\right)^2 \le C_h$ for all j.

(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.

(C6) Let Ω_{Nt}^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_{Nt}^0 are bounded away from zero and infinity.

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. 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 (2015) 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). Conditions (C4) and (C5) are similar to Conditions A2, A5 and A7 of Connor, Matthias and Linton (2012).

Define

$$\Lambda_{Nt}^0 = N^{-1}\sum\nolimits_{i=1}^N E\{p_{it}(0\,|X_i)\,G_i^0(X_i)G_i^0(X_i)^\intercal\}.$$

and

$$\mathbf{\Sigma}_{Nt}^0 = \tau(1-\tau)(\Lambda_{Nt}^0)^{-1}\Omega_{Nt}^0(\Lambda_{Nt}^0)^{-1}. \tag{4.1}$$

The theorem below presents the asymptotic distribution of the final estimator $\widehat{f}_t = \widehat{f}_t^{[i+1]}$ given in Section 3.1, for a given finite number $i \ge 0$. Define

$$\phi_{NT} = \sqrt{K_N/(NT)} + K_N^{3/2}N^{-3/4}\sqrt{\log NT} + K_N^{-r}. \tag{4.2}$$

Let d_{NT} be a sequence satisfying

$$d_{NT} = O(\phi_{NT}). \tag{4.3}$$

Theorem 1. *Assume that Conditions (C1)-(C5) hold, and* $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$, $K_N^{-1}(\log NT)(\log N)^4 = o(1)$ *and* $\phi_{NT} = o(1)$. *Then, for a given* t, *there is a stochastically bounded sequence* $\delta_{N,jt}$ *such that as* $(N, T) \to \infty$,

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

where $\delta_{N,t} = (\delta_{N,jt}, 0 \le j \le J)^\intercal$, d_{NT} *is given in (4.3), and* $\mathbf{I}_{J+1}$ *is the* $(J+1) \times (J+1)$ *identity matrix.*

The next theorem establishes the rate of convergence of the final estimator $\widehat{g}_j(x_j) = \widehat{g}_j^{[i+1]}(x_j)$ given in Section 3.1, for a given finite number $i \ge 0$.

Theorem 2. *Suppose that the same conditions as given in Theorem 1 hold. Then, for each* j, *as* $(N, T) \to \infty$,

$$\left[\int \{\widehat{g}_j(x_j) - g_j^0(x_j)\}^2 dx_j\right]^{1/2} = O_p(\phi_{NT}) + o_p(N^{-1/2}), \tag{4.4}$$

where ϕ_{NT} *is given in (4.2).*

Remark 1: Note that in the asymptotic distribution in Theorem 1, there is a bias term $d_{NT}\delta_{N,t}$ involved. In addition to the order requirements of K_N, N and T given in Theorem 1, we also need $\sqrt{N}\phi_{NT} = o(1)$ in order to let the asymptotic bias be negligible, and as a result we have

$$\sqrt{N}(\mathbf{\Sigma}_{Nt}^0)^{-1/2}(\widehat{f}_t - f_t^0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}_{J+1}), \tag{4.5}$$

and $\left[\int \{\widehat{g}_j(x_j) - g_j^0(x_j)\}^2 dx_j\right]^{1/2} = o_p(N^{-1/2})$. These conditions on the orders of K_N, N and T are equivalent to $\max(N^{1/(2r)}, (\log NT)(\log N)^4) \ll K_N \ll \min(T, N^{1/6}\{\log(NT)\}^{-1/3})$, for $r > 3$. This implies that $N^{1/(2r)} \ll K_N \ll T$.

5 Covariance estimation and hypothesis testing for the factors

In order to construct the confidence interval given in (4.5) we need to estimate Ω_{Nt}^0 and Λ_{Nt}^0, since they are unknown. For estimation of Λ_{Nt}^0, if we use its sample analogue, the conditional density $p_{it}(0 \,|X_i)$ 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

$$\widehat{\Lambda}_{Nt} = (Nh)^{-1} \sum\nolimits_{i=1}^N K\left(\frac{y_{it} - \widehat{f}_{ut} - \sum_{j=1}^J \widehat{g}_j(X_{ji})\widehat{f}_{jt}}{h}\right) \widehat{G}_i(X_i)\widehat{G}_i(X_i)^\intercal, \tag{4.6}$$

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^0_{Nt} 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 (4.2). Then, as $(N, T) \to \infty$, for a given t, we have $||\widehat{\Lambda}_{Nt} - \Lambda^0_{Nt}|| = o_p(1)$.*

Moreover, the exact form of Ω^0_{Nt} defined in Condition (C6) is given by

$$\begin{aligned}\Omega^0_{Nt} &= N^{-1}E\left[\left\{\sum\nolimits_{i=1}^N G^0_i(X_i)(\tau - I(\varepsilon_{it} < 0))\right\}\left\{\sum\nolimits_{i=1}^N G^0_i(X_i)(\tau - I(\varepsilon_{it} < 0))\right\}^\intercal\right] \\ &= \frac{\tau(1-\tau)}{N}\sum\nolimits_{i=1}^N E\{G^0_i(X_i)G^0_i(X_i)^\intercal\} + N^{-1}\sum\nolimits_{i\neq j}^N E(v_{it}v_{jt}^\intercal),\end{aligned}$$

where $v_{it} = G^0_i(X_i)(\tau - I(\varepsilon_{it} < 0))$ for $i = 1, \ldots, N$. To estimate Ω^0_{Nt}, its sample analogue is not consistent. Kernel-based robust estimators that account for 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 of 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^0_{Nt} under the quantile setting. The HAC estimator is given as

$$\widehat{\Omega}_{Nt,M} = \frac{\tau(1-\tau)}{N}\sum\nolimits_{i=1}^N \widehat{G}_i(X_i)\widehat{G}_i(X_i)^\intercal + N^{-1}\sum\nolimits_{i\neq j}^N K^*\left(\frac{i-j}{M}\right)\widehat{v}_{it}\widehat{v}_{jt}^\intercal, \tag{4.7}$$

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 $\widehat{\Omega}_{Nt,M}$ needs that $M \to \infty$ and $M/N \to 0$. The following theorem provides the limiting distribution of $\widehat{\Omega}_{Nt,M=bN}$ 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 $\Omega^0_t = \lim_{N\to\infty}\Omega^0_{Nt}$, and Ω^0_t can be written as $\Omega^0_t = \Upsilon_t\Upsilon_t^\intercal$, where Υ_t is a lower triangular matrix obtained from the Cholesky decomposition of Ω^0_t.

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* $(N,T) \to \infty$, *for a given* t,

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

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* $r \in [0,1]$.

Remark 2: Theorem 4 provides the limiting distribution of $\widehat{\Omega}_{Nt,M=bN}$, although $\widehat{\Omega}_{Nt,M=bN}$ is an inconsistent estimator of Ω_t^0. However, it can be used to construct asymptotically pivotal tests involving f_t^0. Establishing the result in Theorem 4 requires ordering of the indices $\{i, 1 \leq i \leq N\}$. This assumption can be relaxed by allowing s_N indices to be mis-assigned, where s_N satisfies $s_N = o(N^{1/2})$. To see this, we let $\{\pi(i), 1 \leq i \leq N\}$ denote a permutation of the indices $\{i, 1 \leq i \leq N\}$, and assume that $\pi(i) \neq i$ for $i \in \mathcal{S}$ and $\pi(i) = i$ for $i \in \mathcal{S}^c$. Let the cardinality of $\mathcal{S}$ be s_N satisfying $s_N = o(N^{1/2})$. Then denote $\widehat{\Omega}^*_{Nt,M}$ the HAC estimator of Ω^0_{Nt} obtained from the data indexed by $\{\pi(i), 1 \leq i \leq N\}$, where $\widehat{\Omega}^*_{Nt,M}$ is defined in the same way as $\widehat{\Omega}^*_{Nt,M}$ with i and j replaced by $\pi(i)$ and $\pi(j)$. It is straightforward to show that $||\widehat{\Omega}_{Nt,M} - \widehat{\Omega}^*_{Nt,M}|| = O_p(s_N^2/N) = o_p(1)$. Therefore, the result in Theorem 4 follows from the Slutsky's theorem.

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

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

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

$$T_{Nt,b} = \frac{N^{1/2}(R\widehat{f}_t - r)}{\sqrt{R\tau(1-\tau)\widehat{\Lambda}_{Nt}^{-1}\widehat{\Omega}_{Nt,M=bN}\widehat{\Lambda}_{Nt}^{-1}\}^{-1}R^{\intercal}}}. \tag{4.8}$$

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* $h \to 0$, $h^{-1}N^{-1/2} = O(1)$, $\phi_{NT}N^{1/2} = o(1)$ *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* $(N,T) \to \infty$, *for a given* t,

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

If $q = 1$, *then as* $(N,T) \to \infty$, *for a given* t,

$$T_{Nt,b} \xrightarrow{\mathcal{D}} \frac{W_1(1)}{\sqrt{\tau(1-\tau)}\sqrt{\int_0^1 \int_0^1 -\frac{1}{b^2} K^{*\prime\prime}\left(\frac{r-s}{b}\right) B_1(r) B_1(s) dr ds}}. \tag{4.9}$$

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* $\Upsilon_t^* = (R\Lambda_t^{-1}\Omega_t^0\Lambda_t^{-1}R^\intercal)^{1/2}$. *Suppose that the same conditions as given in Theorem 5 hold. Let* $M = bN$ *for* $b \in (0,1]$. *Then under the alternative hypothesis* H_1: $Rf_t^0 = r + cN^{-1/2}$, *as* $(N,T) \to \infty$, *for a given* t,

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

If $q = 1$, *then as* $(N,T) \to \infty$, *for a given* t,

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

Remark 3: If $K^*(x)$ is the Bartlett kernel, then

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

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 Monte Carlo Simulations

In this section, we conduct simulation studies to assess the finite-sample performance of our proposed method.

6.1 Data Generating Processes

We generate the responses from the model:

$$y_{it} = f^0_{ut,\tau} + \sum_{j=1}^{2} g^0_{j,\tau}(X_{ji}) f^0_{jt,\tau} + \varepsilon_{it},$$

for $i = 1, ..., N$ and $t = 1, ..., T$, where $f^0_{t,\tau} = (f^0_{ut,\tau}, f^0_{1t,\tau}, f^0_{2t,\tau})^\top$ and $X_i = (X_{1i}, X_{2i})^\top$. We obtain $\{f^0_{ut,\tau}, 1 \le t \le T\}$ from the multivariate normal distribution with mean $(1.5 + |0.5 - \tau|)\mathbf{1}_T$, where $\mathbf{1}_T$ is the T-dimensional vector of ones, and covariance $\boldsymbol{\Sigma}_f = \{\sigma_{tt'}\} = 0.4^{|t-t'|+2}$, and obtain $\{f^0_{jt,\tau}, 1 \le t \le T\}$ from the same distribution, for $j = 1, 2$. We fix the values of f^0_t, $t = 1, ..., T$, for all simulation replications. We simulate X_{1i} and X_{2i} from $U(-1, 1)$, respectively. Let $g^0_{1,\tau}(x_1) = 0.5(2.5 + 0.5\tau)\cos(\pi x)$ and $g^0_{2,\tau}(x_2) = 0.5(2.5 + 0.5\tau)\sin(\pi x)$. We consider the following setups for the error terms ε_{it}.

Case 1 (Cross sectional dependence): $\boldsymbol{\varepsilon}_{\cdot t} = \{\varepsilon_{it}, 1 \le i \le N\}^\top$ are generated independently from the multivariate t-distribution with the covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{ii'}\} = 0.5^{|i-i'|}$ and 2 degrees of freedom.

Case 2 (Heteroskedasticity): $\varepsilon_{it} = \sigma_i e_{it}$, where σ_i are generated independently from $U(0.5, 1.5)$ and e_{it} are simulated independently from Laplace$(0, 1)$ distribution.

Let $N = 100, 200, 400$ and $T = 20, 40, 100, 200, 400$. All results are based on 500 simulation realizations.

6.2 Results

To evaluate the estimation accuracy of the proposed estimates, we first report the square root of the mean squared error (RMSE) of the estimates $\widehat{f}_{jt}$ and $\widehat{g}_j(\cdot)$, defined as $\{\sum_{j=1}^{J}\sum_{t=1}^{T}(\widehat{f}_{jt} - f^0_{jt})^2/(JT)\}^{1/2}$ and $[\sum_{j=1}^{J}\sum_{i=1}^{N}\{\widehat{g}_j(X_{ji}) - g^0_j(X_{ji})\}^2/(JN)]^{1/2}$, respectively, where $\widehat{f}_{jt}$ and $\widehat{g}_j(\cdot)$ are generic notations for estimators of f_{jt} and $g_j(\cdot)$. We present the RMSE values for three different estimates obtained from the algorithm described in Section 3.1: (1) $\widehat{f}_t = \widehat{f}^{[1]}_t$ and $\widehat{g}_j = \widehat{g}^{[1]}_j$ (the algorithm is stopped at the 1st step); (2) $\widehat{f}_t = \widehat{f}^{[2]}_t$ and $\widehat{g}_j = \widehat{g}^{[2]}_j$ (the algorithm is stopped at the 2nd step); (3) $\widehat{f}_t = \widehat{f}^{[i]}_t$ and $\widehat{g}_j = \widehat{g}^{[i]}_j$ given that the algorithm converges at the i^{th} step (meets the stopping rule (1) given in Section 3.1).The RMSEs of these three estimates are denoted as RMSE_1, RMSE_2 and RMSE_c, respectively. We use B-splines of order $m = 4$ for estimating the loading functions, and the number of interior knots for the B-splines are selected by the BIC given in Section 7. We consider three quantiles $\tau = 0.3, 0.5$ and 0.7. Tables 1-3 report the median value of the RMSEs for the estimates of the factors and their loading functions obtained from the 500 simulation replications at $\tau = 0.3, 0.5$ and 0.7, respectively. We observe that the values of RMSE_2 and RMSE_c are almost identical for all cases. This indicates that the estimates at the 2^{nd} step behave similarly to the estimates at convergence. In general, the RMSE_1 values are close to the corresponding values of RMSE_2 and RMSE_c. However, for the estimates of the factors, the RMSE_1 values are obviously larger than the RMSE_2 and RMSE_c values for some cases; see the results for Case 1 in Tables 2 and 3. As a result, we recommend to use the estimates at the i^{th} step, where i is finite and $i \geq 2$, or the estimates at convergence. Moreover, we see that the RMSE values decrease as N increases. This corroborates the asymptotic results obtained in Section 4, and the estimates perform well for all the T values considered in our simulations.

Table 1: The median value of the RMSEs based on the 500 simulation replications at $\tau = 0.3$.

	factors			factor loading functions		
(N,T)	RMSE_1	RMSE_2	RMSE_c	RMSE_1	RMSE_2	RMSE_c
Case 1						
$(100,20)$	0.311	0.313	0.313	0.124	0.124	0.124
$(200,20)$	0.237	0.240	0.240	0.100	0.100	0.100
$(400,20)$	0.174	0.176	0.176	0.082	0.082	0.082
$(100,40)$	0.314	0.318	0.318	0.118	0.118	0.118
$(200,40)$	0.240	0.243	0.243	0.097	0.097	0.097
$(400,40)$	0.179	0.182	0.182	0.079	0.079	0.079
$(100,100)$	0.322	0.325	0.325	0.114	0.114	0.114
$(200,100)$	0.243	0.248	0.248	0.095	0.095	0.095
$(400,100)$	0.184	0.186	0.186	0.078	0.078	0.078
$(100,200)$	0.337	0.340	0.340	0.113	0.113	0.113
$(200,200)$	0.249	0.253	0.253	0.095	0.095	0.095
$(400,200)$	0.185	0.188	0.188	0.078	0.078	0.078
$(100,400)$	0.354	0.358	0.358	0.113	0.113	0.113
$(200,400)$	0.262	0.264	0.264	0.094	0.094	0.094
$(400,400)$	0.190	0.192	0.192	0.078	0.078	0.078
Case 2						
$(100,20)$	0.120	0.122	0.122	0.105	0.105	0.105
$(200,20)$	0.084	0.089	0.089	0.080	0.080	0.080
$(400,20)$	0.067	0.070	0.070	0.055	0.055	0.055
$(100,40)$	0.120	0.124	0.124	0.100	0.100	0.100
$(200,40)$	0.088	0.092	0.092	0.079	0.079	0.079
$(400,40)$	0.067	0.070	0.070	0.054	0.054	0.054
$(100,100)$	0.125	0.129	0.129	0.098	0.098	0.098
$(200,100)$	0.082	0.088	0.088	0.077	0.077	0.077
$(400,100)$	0.065	0.069	0.069	0.053	0.053	0.053
$(100,200)$	0.120	0.122	0.122	0.098	0.098	0.098
$(200,200)$	0.085	0.090	0.090	0.077	0.077	0.077
$(400,200)$	0.066	0.069	0.069	0.053	0.053	0.053
$(100,400)$	0.122	0.128	0.128	0.097	0.097	0.097
$(200,400)$	0.087	0.092	0.092	0.077	0.077	0.077
$(400,400)$	0.065	0.068	0.068	0.053	0.053	0.053

Next, we let $f_{t,\tau} = f^0_{t,\tau} + cN^{-1/2}\mathbf{1}_{J+1}$ for a given c, and simulate the responses from the model

Table 2: The median value of the RMSEs based on the 500 simulation replications at $\tau = 0.5$.

	factors			factor loading functions		
(N, T)	RMSE_1	RMSE_2	RMSE_c	RMSE_1	RMSE_2	RMSE_c
Case 1						
$(100, 20)$	0.300	0.287	0.287	0.114	0.114	0.114
$(200, 20)$	0.204	0.200	0.200	0.093	0.093	0.093
$(400, 20)$	0.166	0.164	0.164	0.079	0.079	0.079
$(100, 40)$	0.302	0.292	0.292	0.113	0.113	0.113
$(200, 40)$	0.218	0.216	0.216	0.092	0.092	0.092
$(400, 40)$	0.171	0.170	0.170	0.078	0.078	0.078
$(100, 100)$	0.376	0.308	0.308	0.115	0.115	0.115
$(200, 100)$	0.271	0.230	0.230	0.094	0.094	0.094
$(400, 100)$	0.200	0.176	0.176	0.079	0.079	0.079
$(100, 200)$	0.380	0.316	0.316	0.114	0.114	0.114
$(200, 200)$	0.276	0.238	0.238	0.093	0.093	0.093
$(400, 200)$	0.210	0.184	0.184	0.078	0.078	0.078
$(100, 400)$	0.388	0.324	0.324	0.113	0.113	0.113
$(200, 400)$	0.290	0.250	0.250	0.093	0.093	0.093
$(400, 400)$	0.215	0.189	0.189	0.078	0.078	0.078
Case 2						
$(100, 20)$	0.126	0.127	0.127	0.104	0.104	0.104
$(200, 20)$	0.086	0.091	0.091	0.080	0.080	0.080
$(400, 20)$	0.066	0.069	0.069	0.055	0.055	0.055
$(100, 40)$	0.122	0.125	0.125	0.100	0.100	0.100
$(200, 40)$	0.085	0.088	0.088	0.078	0.078	0.078
$(400, 40)$	0.063	0.068	0.068	0.053	0.053	0.053
$(100, 100)$	0.122	0.126	0.126	0.098	0.098	0.098
$(200, 100)$	0.085	0.090	0.090	0.077	0.077	0.077
$(400, 100)$	0.064	0.068	0.068	0.053	0.053	0.053
$(100, 200)$	0.120	0.123	0.123	0.098	0.098	0.098
$(200, 200)$	0.083	0.088	0.088	0.077	0.077	0.077
$(400, 200)$	0.064	0.067	0.067	0.053	0.053	0.053
$(100, 400)$	0.120	0.124	0.124	0.097	0.097	0.097
$(200, 400)$	0.083	0.088	0.088	0.077	0.077	0.077
$(400, 400)$	0.064	0.067	0.067	0.053	0.053	0.053

Table 3: The median value of the RMSEs based on the 500 simulation replications at $\tau = 0.7$.

	factors			factor loading functions		
(N, T)	RMSE_1	RMSE_2	RMSE_c	RMSE_1	RMSE_2	RMSE_c
Case 1						
$(100, 20)$	0.311	0.285	0.285	0.106	0.106	0.106
$(200, 20)$	0.223	0.205	0.205	0.081	0.081	0.081
$(400, 20)$	0.155	0.144	0.144	0.051	0.051	0.051
$(100, 40)$	0.319	0.290	0.290	0.098	0.098	0.098
$(200, 40)$	0.239	0.215	0.215	0.078	0.078	0.078
$(400, 40)$	0.155	0.148	0.148	0.047	0.047	0.047
$(100, 100)$	0.344	0.300	0.299	0.096	0.096	0.096
$(200, 100)$	0.247	0.224	0.224	0.074	0.074	0.074
$(400, 100)$	0.161	0.150	0.150	0.046	0.046	0.046
$(100, 200)$	0.361	0.312	0.312	0.094	0.094	0.094
$(200, 200)$	0.249	0.227	0.227	0.074	0.074	0.074
$(400, 200)$	0.169	0.160	0.160	0.045	0.045	0.045
$(100, 400)$	0.371	0.326	0.326	0.094	0.094	0.094
$(200, 400)$	0.263	0.241	0.241	0.074	0.074	0.074
$(400, 400)$	0.175	0.169	0.169	0.045	0.045	0.045
Case 2						
$(100, 20)$	0.120	0.122	0.122	0.105	0.105	0.105
$(200, 20)$	0.086	0.090	0.090	0.080	0.080	0.080
$(400, 20)$	0.064	0.068	0.068	0.055	0.055	0.055
$(100, 40)$	0.122	0.125	0.125	0.100	0.100	0.100
$(200, 40)$	0.085	0.090	0.090	0.078	0.078	0.078
$(400, 40)$	0.065	0.069	0.069	0.054	0.054	0.054
$(100, 100)$	0.120	0.122	0.122	0.098	0.098	0.098
$(200, 100)$	0.087	0.092	0.092	0.077	0.077	0.077
$(400, 100)$	0.065	0.069	0.069	0.053	0.053	0.053
$(100, 200)$	0.120	0.122	0.122	0.098	0.098	0.098
$(200, 200)$	0.086	0.091	0.091	0.077	0.077	0.077
$(400, 200)$	0.065	0.069	0.069	0.053	0.053	0.053
$(100, 400)$	0.125	0.129	0.129	0.097	0.097	0.097
$(200, 400)$	0.087	0.092	0.092	0.077	0.077	0.077
$(400, 400)$	0.065	0.069	0.069	0.053	0.053	0.053

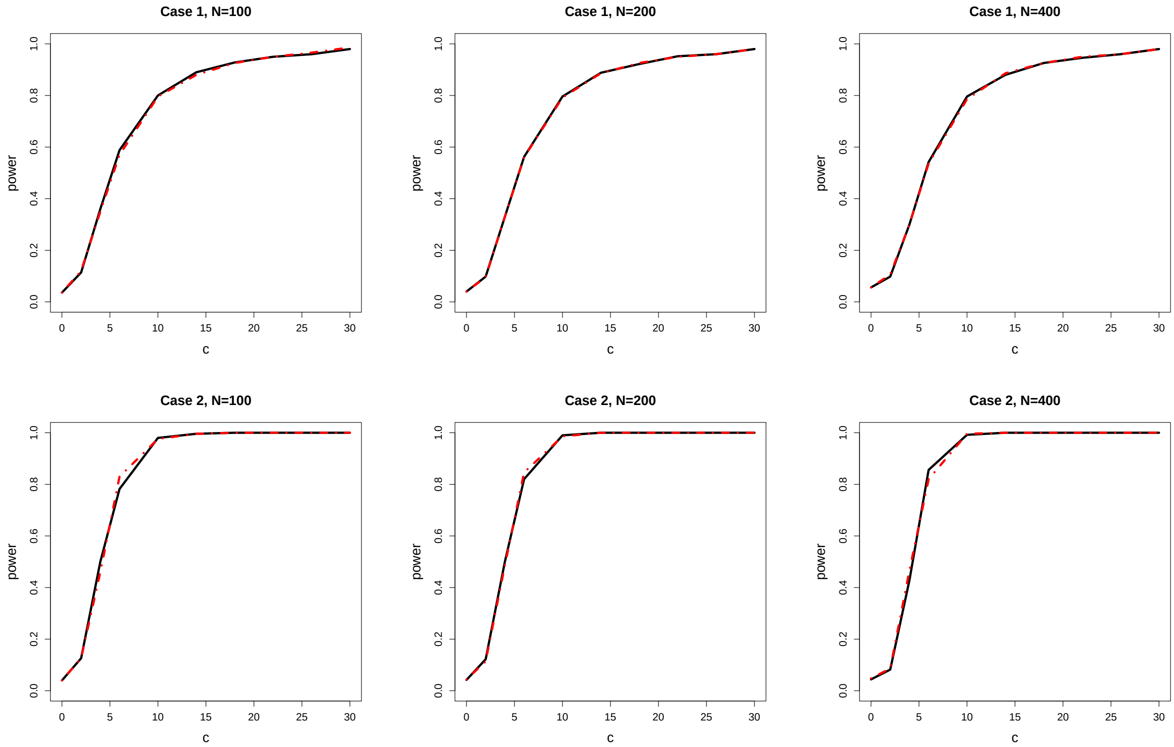
(2.2) using these factors and the same loading functions $g^0_{j,\tau}(\cdot)$ given in Section 6.1. The error terms ε_{it} are generated from the two cases described in Section 6.1. We consider the hypotheses: H_0: $f_{jt,\tau} = f^0_{jt,\tau}$ versus H_1: $f_{jt,\tau} \neq f^0_{jt,\tau}$ for $j = 1, 2$, respectively. Based on the 500 simulation realizations, we use the proposed t-type statistic $T_{Nt,b}$ given in (4.8) to obtain the empirical power= $\sum_{s=1}^{500}\{I(T_{Nt,b,s} > \mathcal{T}_{1-\alpha/2}) + I(T_{Nt,b,s} < \mathcal{T}_{\alpha/2})\}/500$ at the significance level $\alpha = 0.05$, where $T_{Nt,b,s}$ is the value of the s^{th} replicate of $T_{Nt,b}$, and $\mathcal{T}_{\alpha/2}$ and $\mathcal{T}_{1-\alpha/2}$ are the $(\alpha/2)^{\text{th}}$ and $(1-\alpha/2)^{\text{th}}$ quantiles of the null limiting distribution given in (4.9). The estimates of the factors and their loading functions are obtained by using the algorithm with the stopping rule (1) given in Section 3.1. For calculation of $\widehat{\Lambda}_{Nt}$ and $\widehat{\Omega}_{Nt,M=bN}$ in (4.8), we let $h = \kappa N^{-1/5}$ with $\kappa = 0.5, 1, 1.5$ and $b = 0.2, 0.4, 0.6$, respectively. The results for these κ and b values are similar, so we choose to report the results for $\kappa = 0.5$ and $b = 0.6$. Figure 1 displays the average value of the empirical powers over all t's versus the c values for $N = 100, 200, 400$, and $T = 20$ (black solid line) and $T = 400$ (red dashed line) at $\tau = 0.5$. The plots of the empirical powers at $\tau = 0.3$, 0.5 and 0.7 look similar, so we only report the plots at $\tau = 0.5$ for saving spaces. The c values range from 0 to 30. We see that the Type I error rates (the powers at $c = 0$) are close to the nominal significance level 0.05 for all cases. Moreover, the empirical size of power increases rapidly to 1 as c increases, and the proposed test has similar performance at $T = 20$ and $T = 400$.

7 Application

In a series of important papers, Fama and French (hereafter denoted FF), demonstrated that there have been large return premia associated with size and value, which are observable characteristics of stocks. They 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. FF (1993) used a simple portfolio sorting approach to estimating their factor model. Connor, Hagmann, and Linton (2012) used kernel-based semiparametric regression methodology to capture the same phenomenon.

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 347 stocks. We consider the same four characteristic variables as given in Connor, Hagmann and Linton (2012), and Fan, Liao and Wang (2016), which are size, value, momentum and volatility. Connor, Hagmann

Figure 1: The plots of the average value of the empirical powers over all t's versus the c values for $T = 20$ (black solid line) and $T = 400$ (red dashed line) at $\tau = 0.5$.



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).

It is commonly accepted based on empirical evidence that financial asset returns are heavy-tailed (Bradley and Taqqu, 2003). One attractive feature of the quantile factor model (2.1) is that it offers a parsimonious way of characterizing the entire conditional distribution, and moreover it can be more robust to heavy tails and outliers than mean factor models. To this end, we fit model (2.1) for each year, so that there are $T = 251$ observations. By taking the same strategy as He and Shi (1996), we select the number of interior knots L_N by minimizing the Bayesian information criterion (BIC) given as

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

For comparison, we fit the quantile model given in Connor, Hagmann, and Linton (2012) through mean regression, in which we use the B-splines to approximate the unknown loading functions. We use the method given in Koenker and Machado (1999) to find the pseudo$-R^2$ for each stock at different quantiles. Then we obtain the average pseudo$-R^2$ among the 347 stocks, which is 0.441, 0.422, 0.428 at quantiles $\tau = 0.2, 0.5, 0.8$, respectively. The average R^2 from the mean regression is 0.312. This indicates that a higher percentage of variability of asset returns may be explained by the quantile model compared to the mean factor model.

Next, we test significance of the factors using the proposed t-type statistic given in Theorem 5. To use the asymptotic distribution given in Theorem 5 for statistical inference, we need to estimate the covariance matrix. For the estimator $\widehat{\Lambda}_{Nt}$ given in (4.6), the optimal order for the bandwidth h is in the order of $N^{-1/5}$. 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 (4.7), we use different values for b, and use the Bartlett kernel as suggested in Kiefer and Vogelsang (2005). We let $\kappa = 0.5, 1, 1.5$ and $b = 0.2, 0.4, 0.6$, respectively, for calculation of $\widehat{\Lambda}_{Nt}$ and $\widehat{\Omega}_{Nt,M=bN}$. For obtaining the robust estimator $\widehat{\Omega}_{Nt,M=bN}$, the data need to be ordered across i. We consider two different orderings. First, we take the same strategy as Lee and Robinson (2016) by ordering the data according to firm size, since firms of similar size may be subject to similar shocks. Second, we use the information of the four explanatory variables by ordering the data according to the first principal component of the covariate matrix. We then test for the statistical significance of each factor at each time point based on the proposed t-type statistic. 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. Based on the two different ordering strategies, Tables 1 and 2, respectively, show the annualized standard deviations of the factor returns, the percentage of significant t-type statistics for each factor, and the median p-value at $\tau = 0.5$ for the year of 2012. We obtain similar results for other years. We can see that the results are consistent for different values of κ and b and for the two different orderings of the data. Moreover, all five factors are statistically significant with the median p-value smaller than 0.05.

For comparison, we also conduct tests for significance of each factor via fitting the mean factor model and using the asymptotic distribution of the estimated factors given in Theorem 1 of Connor, Hagmann, and Linton (2012). Table 6 shows the percentage of significant statistics for each factor and the median p-value from fitting the quantile regression and the mean regression (mean), respectively. We let $\kappa = 1$ and $b = 0.2$ for calculation of $\widehat{\Lambda}_{Nt}$ and $\widehat{\Omega}_{Nt,M=bN}$ and order the data according to the first principal component of the covariate matrix. We see that the quantile regression identifies more significant factors across time than the mean regression does.

Lastly, we plot the four estimated loading functions and the 95% pointwise confidence intervals at different quantiles $\tau = 0.2$, 0.5 and 0.8 for the year of 2012. The plots of the estimated functions have similar patterns from other years. The pointwise confidence intervals are obtained from a wild bootstrap procedure (Mammen, 1993). Figure 2 shows the plots of the estimated loading functions and the pointwise confidence intervals at quantile $\tau = 0.5$. The solid black lines are the estimated

Table 4: Factor return statistics at $\tau = 0.5$ for the year of 2012 when the data are ordered according to the firm size.

(κ, b)		Intercept	Size	Value	Momentum	Volatility
	Annualized volatility	0.024	0.023	0.025	0.025	0.029
$(0.5, 0.2)$	% Periods significant	92.43	65.34	62.95	64.54	74.10
	Overall p-value	< 0.001	0.010	0.016	0.009	0.001
	Annualized volatility	0.020	0.020	0.022	0.022	0.026
$(0.5, 0.4)$	% Periods significant	91.63	58.17	57.20	58.17	66.93
	Overall p-value	< 0.001	0.020	0.019	0.020	0.010
	Annualized volatility	0.018	0.019	0.019	0.020	0.023
$(0.5, 0.6)$	% Periods significant	90.84	55.78	56.40	55.38	66.93
	Overal p-value	< 0.001	0.032	0.028	0.023	0.006
	Annualized volatility	0.026	0.026	0.027	0.027	0.032
$(1.0, 0.2)$	% Periods significant	92.03	62.95	63.60	62.15	71.31
	Overall p-value	< 0.001	0.014	0.019	0.011	0.002
	Annualized volatility	0.022	0.023	0.023	0.024	0.028
$(1.0, 0.4)$	% Periods significant	90.44	55.20	56.40	55.98	65.74
	Overall p-value	< 0.001	0.036	0.030	0.033	0.011
	Annualized volatility	0.019	0.022	0.021	0.021	0.025
$(1.0, 0.6)$	% Periods significant	89.24	56.20	55.40	58.80	62.95
	Overall p-value	< 0.001	0.032	0.032	0.026	0.016
	Annualized volatility	0.027	0.028	0.029	0.029	0.034
$(1.5, 0.2)$	% Periods significant	92.03	59.76	55.38	61.75	70.12
	Overall p-value	< 0.001	0.021	0.032	0.015	0.003
	Annualized volatility	0.023	0.025	0.025	0.026	0.031
$(1.5, 0.4)$	% Periods significant	90.44	56.57	55.94	55.94	63.75
	Overall p-value	< 0.001	0.030	0.030	0.036	0.014
	Annualized volatility	0.020	0.019	0.022	0.022	0.026
$(1.5, 0.6)$	% Periods significant	88.44	58.14	56.80	56.00	61.75
	Overall p-value	< 0.001	0.027	0.028	0.024	0.018

Table 5: Factor return statistics at $\tau = 0.5$ for the year of 2012 when the data are ordered according to the first principal component of the covariate matrix.

(κ, b)		Intercept	Size	Value	Momentum	Volatility
	Annualized volatility	0.023	0.027	0.025	0.025	0.027
$(0.5, 0.2)$	% Periods significant	94.02	62.15	62.55	67.73	75.30
	Overall p-value	< 0.001	0.023	0.018	0.011	< 0.001
	Annualized volatility	0.019	0.024	0.022	0.021	0.023
$(0.5, 0.4)$	% Periods significant	92.43	57.60	54.20	58.96	70.92
	Overall p-value	< 0.001	0.023	0.032	0.019	0.001
	Annualized volatility	0.016	0.021	0.020	0.019	0.020
$(0.5, 0.6)$	% Periods significant	92.83	55.60	56.40	61.60	71.31
	Overal p-value	< 0.001	0.028	0.028	0.018	0.004
	Annualized volatility	0.025	0.032	0.027	0.027	0.030
$(1.0, 0.2)$	% Periods significant	93.23	56.80	60.40	64.80	74.30
	Overall p-value	< 0.001	0.033	0.021	0.016	0.001
	Annualized volatility	0.020	0.026	0.024	0.024	0.025
$(1.0, 0.4)$	% Periods significant	92.03	54.80	56.20	59.60	71.20
	Overall p-value	< 0.001	0.030	0.030	0.019	0.002
	Annualized volatility	0.016	0.024	0.022	0.022	0.022
$(1.0, 0.6)$	% Periods significant	92.80	56.20	55.40	56.80	68.80
	Overall p-value	< 0.001	0.027	0.031	0.029	0.002
	Annualized volatility	0.027	0.030	0.029	0.028	0.032
$(1.5, 0.2)$	% Periods significant	92.03	56.00	54.40	68.00	74.00
	Overall p-value	< 0.001	0.033	0.032	0.013	0.002
	Annualized volatility	0.021	0.028	0.026	0.026	0.026
$(1.5, 0.4)$	% Periods significant	92.03	56.60	55.90	55.20	68.00
	Overall p-value	< 0.001	0.028	0.028	0.030	0.002
	Annualized volatility	0.018	0.025	0.024	0.023	0.024
$(1.5, 0.6)$	% Periods significant	92.03	58.10	54.80	56.00	67.60
	Overall p-value	< 0.001	0.027	0.030	0.029	0.003

Table 6: P-value statistics for testing significance of factors from quantile regression at $\tau = 0.2$, 0.5, 0.8 and from mean regression (mean) for the year of 2012 when the data are ordered according to the first principal component of the covariate matrix.

		Intercept	Size	Value	Momentum	Volatility
$\tau = 0.2$	% Periods significant	97.20	63.20	62.80	58.00	83.20
	Overall p-value	< 0.001	0.015	0.011	0.022	< 0.001
$\tau = 0.5$	% Periods significant	93.23	56.80	60.40	64.80	74.30
	Overall p-value	< 0.001	0.033	0.021	0.016	0.001
$\tau = 0.8$	% Periods significant	97.20	66.40	65.60	71.20	79.60
	Overal p-value	< 0.001	0.004	0.010	0.008	< 0.001
mean	% Periods significant	0.840	0.152	0.312	0.264	0.456
	Overall p-value	< 0.001	0.314	0.314	0.173	0.071

loading functions from the quantile regression, the dashed blue lines are the confidence intervals and the dotted red lines are the estimated loading functions from the mean regression. For the size, value and momentum characteristics, the estimated functions show a clear nonlinear pattern. This indicates that the effects of these characteristics on the stock returns change with their values. For instance, at $\tau = 0.5$, both small and big sizes of firms tend to have a strong effect on stock returns. However, the effect of size becomes insignificant when its value falls into certain range. For other three characteristics, their effects are strong for most of their values. In Figures 3 and 4, we also plot the estimated loading functions and the 95% confidence intervals at quantiles $\tau = 0.2$ and 0.8, respectively. We see that the four characteristics can have different effects on the stock returns at different quantiles. For example, at $\tau = 0.2$ and 0.8, the effect of size fluctuates around zero, and it suddenly becomes strong after the value of size exceeds certain value. This patten is quite different from what we observe for the size characteristic at $\tau = 0.5$ shown in Figure 2. The results indicate that the effect of small firm sizes on stock return becomes weaker from the middle to tails of its distribution.

8 Conclusions and discussion

Our semiparametric quantile factor models are motivated by the classical factor models (Bai and Ng, 2002) and the characteristic-based factor models (Connor et al., 2012), in which the factor betas are time-invariant, i.e., there is no structural change in the factor betas over time. The four characteristics considered in Connor et al. (2012) and Fan et al. (2016) are used as the time-invariant baseline covariates in the loading functions, and they are calculated using the data right before the data analyzing window. In practice, the firm characteristics can be varying over time. It is of interest to consider an extended model which incorporates the time-varying characteristics. This model can be given as $Q_{y_{it}}(\tau|X_{it}) = f_{ut,\tau} + \sum_{j=1}^{J} g_{j,\tau}(X_{jit}) f_{jt,\tau}$, where X_{jit} are observable time-dependent characteristics. Further investigations are needed for developing the estimating methods for this model and the associated statistical properties. We will consider this interesting yet challenging problem as a future research topic to explore.

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

$$\underset{T\to\infty}{\mathrm{p}\lim} \frac{1}{T} \sum_{t=1}^{T} f_{jt}^{0} \neq 0 \tag{8.1}$$

for $j = 1, \ldots, J$. For the factors in our application this is quite a standard assumption, but in some

Figure 2: The plots of the estimated loading functions from quantile regression (solid black lines) at $\tau = 0.5$, the 95% pointwise confidence intervals (dashed blue lines) and the estimated loading functions from mean regression (dotted red lines).

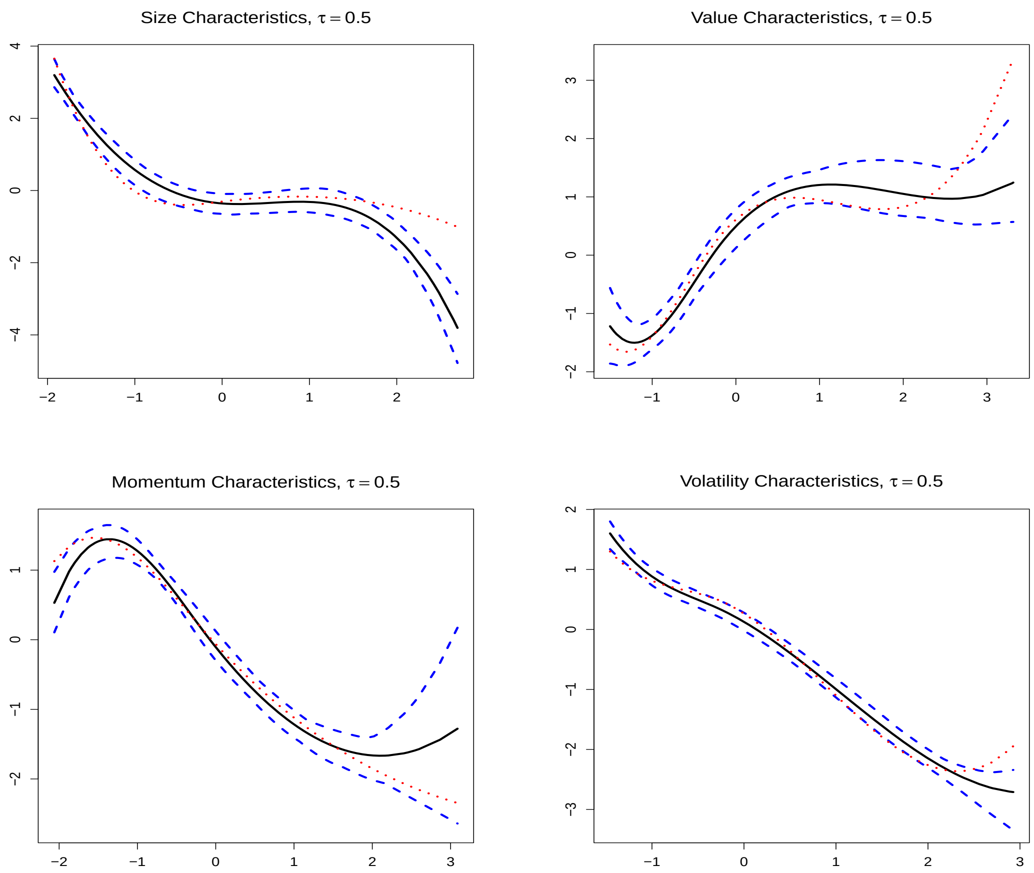


Figure 3: The plots of the estimated loading functions from quantile regression (solid black lines) and the 95% pointwise confidence intervals (dashed blue lines) at $\tau = 0.2$.

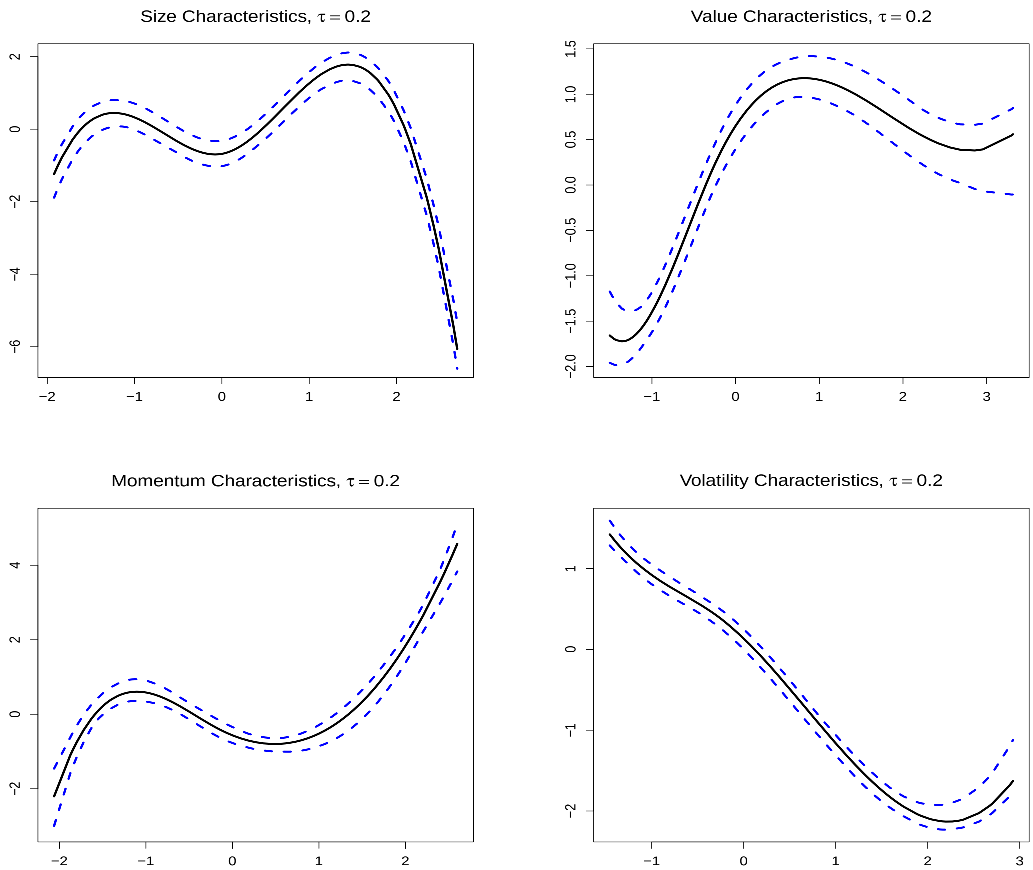
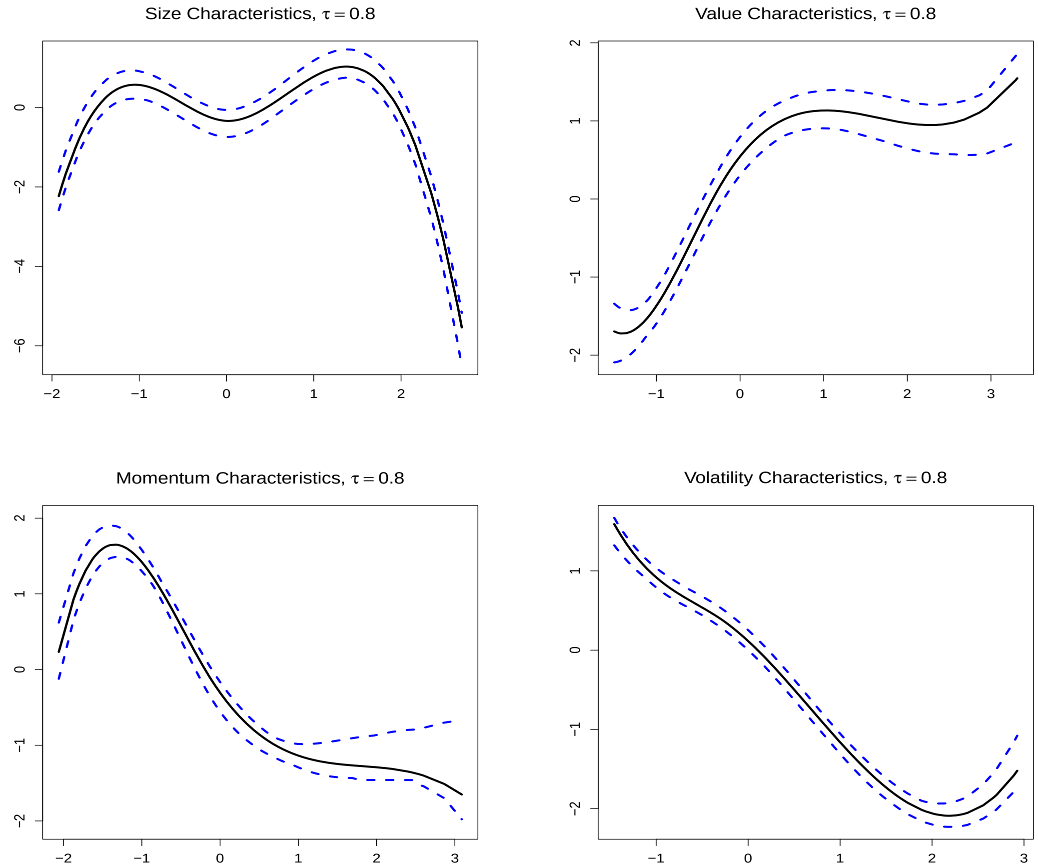


Figure 4: The plots of the estimated loading functions from quantile regression (solid black lines) and the 95% pointwise confidence intervals (dashed blue lines) at $\tau = 0.8$.



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 cannot identify $g_j^0(x_j)$. We outline below a test of the hypothesis (8.1) based on the unstructured additive quantile regression model (2.3). A more limited objective is to test whether for a given time period t, $f_{jt} = 0$.

We are interested in testing the hypothesis that

$$H_{0_{A_j}} : \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} h_{jt}(x_j) = 0 \text{ for all } x_j, \tag{8.2}$$

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 $E\{\mu_j(X_{ji})^2\} > 0$. 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, which is a subset of $\{1, 2, \ldots, J\}$. These are tests of the presence of a factor.

We let

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

where $\widehat{h}_{jt}(\cdot)$ is an estimator of the additive component function $h_{jt}(\cdot)$ from the quantile additive model at time t, while $a_{N,T}$ and $s_{N,T}$ are constants to be determined. Under the null hypothesis (8.2) we may show that

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

while under the alternative we have $\widehat{\tau}_{j,n,T} \to \infty$ with probability approaching one. To ensure that $\widehat{\tau}_{j,n,T}$ has an asymptotic distribution, we may need a two-step estimator for the additive functions $h_{jt}(\cdot)$ as given in Horowitz and Mammen (2011) or Ma and Yang (2011). This interesting and challenging technical problem deserves further investigation, and it can be a good future research topic.

9 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 $\mathbf{A}$, 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 estimators $\widehat{g}_j^{[0]}(x_j)$ of $g_j^0(x_j)$. The following proposition gives the convergence rate of $\widehat{g}_j^{[0]}(x_j)$ 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* $1 \le j \le J$, *as* $(N,T) \to \infty$,

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

9.1 Proof of Proposition 1

According to the result on page 149 of de Boor (2001) and Lemma 2 in Xue and Yang (2006), for h_{jt}^0 satisfying the smoothness condition given in (C2), there exists $\boldsymbol{\theta}_{jt}^0 \in \mathbb{R}^{K_n}$ such that $h_{jt}^0(x_j) = \widetilde{h}_{jt}^0(x_j) + b_{jt}(x_j)$

$$\widetilde{h}_{jt}^0(x_j) = B_j(x_j)^\intercal \boldsymbol{\theta}_{jt}^0 \text{ and } \sup_{j,t} \sup_{x_j \in [a,b]} |b_{jt}(x_j)| = O(K_N^{-r}). \tag{A.1}$$

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

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

where $B(x) = \{B_1(x_1)^\intercal, \ldots, B_J(x_J)^\intercal\}^\intercal$ and $\boldsymbol{\theta}_t^0 = (\boldsymbol{\theta}_{1t}^{0\intercal}, \ldots, \boldsymbol{\theta}_{Jt}^{0\intercal})^\intercal$. Then by (A.1), we have

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

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

$$\mathbb{B}(x) = [\text{diag}\{1, B_1(x_1)^\intercal, \ldots, B_J(x_J)^\intercal\}]_{(1+J)\times(1+JK_N)}, \tag{A.2}$$

$\widetilde{h}_t(x) = \{\widetilde{h}_{jt}(x_j), 1 \le j \le J\}^\intercal$, and $\widetilde{h}_{jt}(\cdot)$ are the estimators given in Section 3.2. 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* $\sup_{i \ge 1} |\xi_i| \le M$. *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 Merlevède, Peligrad and Rio (2009) when the sequence $\{\xi_i\}$ has the α-mixing coefficient satisfying $\alpha(k) \leq \exp(-2ck)$ for some $c > 0$. Thus, this result also holds for the sequence having the ϕ-mixing coefficient satisfying $\phi(k) \leq \exp(-2ck)$, since $\alpha(k) \leq \phi(k) \leq \exp(-2ck)$. ∎

Denote $B(X_i) = \{B_1(X_{1i})^\intercal, \ldots, B_J(X_{Ji})^\intercal\}^\intercal$ and $Z_i = [\{1, B(X_i)^\intercal\}^\intercal]_{(1+JK_N)\times 1}$. 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

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

where $F_{it}(\varepsilon|X_i) = P(\varepsilon_{it} \leq \varepsilon|X_i)$, and $\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t) = G_{tN,i}(\boldsymbol{\vartheta}_t) - G_{tN,i}^*(\boldsymbol{\vartheta}_t)$. Let $d(N) = (1 + JK_N)$.

Let $\Psi_{Nt} = N^{-1}\sum_{i=1}^N p_{it}\left(0\,|X_i\right) Z_i Z_i^\intercal$. 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 $0 < C_1 \leq C_2 < \infty$ such that

$$C_1 \leq \lambda_{\min}(\Psi_{Nt}) \leq \lambda_{\max}(\Psi_{Nt}) \leq C_2, \tag{A.3}$$

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

Next lemma presents the Bahadur representation for $\widetilde{\boldsymbol{\vartheta}}_t = (\widetilde{h}_{ut}, \widetilde{\boldsymbol{\theta}}_t^\intercal)^\intercal$ using the results in Lemmas **A.1-A.3** given in the Supplemental Materials.

Lemma 2. *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* $K_N^{-r+1}(\log T) = o(1)$,

$$\widetilde{\boldsymbol{\vartheta}}_t - \boldsymbol{\vartheta}_t^0 = D_{Nt,1} + D_{Nt,2} + R_{Nt}, \tag{A.4}$$

where

$$D_{Nt,1} = \Psi_{Nt}^{-1}\left[N^{-1}\sum\nolimits_{i=1}^N Z_i(\tau - I(\varepsilon_{it} < 0))\right], \tag{A.5}$$

$$D_{Nt,2} = \Psi_{Nt}^{-1}\left[N^{-1}\sum\nolimits_{i=1}^N Z_i\{p_{it}\left(0\,|X_i\right)\sum\nolimits_{j=1}^J b_{jt}(X_{ji})\}\right],$$

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

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

Proof. By Lemma **A.3** in the Supplemental Materials, we have

$$\widetilde{\boldsymbol{\vartheta}}_t - \boldsymbol{\vartheta}_t^0 = N^{-1}\Psi_{Nt}^{-1}\sum\nolimits_{i=1}^{N} p_{it}\,(0\,|X_i\,)\,Z_i b_t(X_i) - N^{-1}\Psi_{Nt}^{-1}\sum\nolimits_{i=1}^{N} G_{tN,i}^{*}(\widetilde{\boldsymbol{\vartheta}}_t) + R_{Nt}^{*}.$$

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,

$$\widetilde{\boldsymbol{\vartheta}}_t - \boldsymbol{\vartheta}_t^0 = \Psi_{Nt}^{-1}N^{-1}\sum\nolimits_{i=1}^{N}\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0) + \Psi_{Nt}^{-1}N^{-1}\sum\nolimits_{i=1}^{N} p_{it}\,(0\,|X_i\,)\,Z_i b_t(X_i) + R_{Nt}^{**}, \qquad (A.6)$$

where

$$R_{Nt}^{**} = -\Psi_{Nt}^{-1}N^{-1}\sum\nolimits_{i=1}^{N} G_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t) + \Psi_{Nt}^{-1}N^{-1}\sum\nolimits_{i=1}^{N}[\widetilde{G}_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t) - \widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)] + R_{Nt}^{*}. \qquad (A.7)$$

By Lemmas **A.1** and **A.2** in the Supplemental Materials and (A.3), 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,

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

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

$$
\begin{aligned}
&E[N^{-1}\sum\nolimits_{i=1}^{N}\{\widetilde{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)-\overline{G}_{tN,i\ell}(\boldsymbol{\vartheta}_t^0)\}]^2\\
&\le N^{-1}CK_N^{-r}+N^{-2}\sum\nolimits_{i\neq i'}C'K_1e^{-\lambda_1|i'-i|}K_N^{-r}\\
&\le CN^{-1}K_N^{-r}+C'K_1N^{-2}N(1-e^{-\lambda_1/2})^{-1}K_N^{-r}\le C''N^{-1}K_N^{-r},
\end{aligned}
$$

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

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

Therefore, by the Bernstein's inequality in Lemma 1 and the union bound of probability, following the same procedure as the proof for Lemma **A.1** given in the Supplemental Materials, we have

$$
\sup_{1\le t\le T}||N^{-1}\sum\nolimits_{i=1}^{N}\{\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)-\overline{G}_{tN,i}(\boldsymbol{\vartheta}_t^0)\}||=O_p(N^{-1/2}K_N^{-r/2+1/2}\sqrt{\log K_NT}). \tag{A.8}
$$

Therefore, by (A.6), (A.7) and (A.8), we have $\widetilde{\boldsymbol{\vartheta}}_t-\boldsymbol{\vartheta}_t^0=D_{Nt,1}+D_{Nt,2}+R_{Nt}$, where

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

∎

Proof of Proposition 1. Let 1_l be the $(J+1)\times 1$ vector with the l^{th} element as "1" and other elements as "0". By (A.4) in Lemma 2, we have

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

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

and

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

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

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

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

$$\widetilde{h}_{jt}(x_j)-h_{jt}^0(x_j)=1_{j+1}^{\intercal}\mathbb{B}(x)D_{Nt,1}+\eta_{N,jt}(x_j), \tag{A.9}$$

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

$$\sup_{1\le t\le T}[N^{-1}\sum\nolimits_{i=1}^N\{\eta_{N,jt}(X_{ji})\}^2]^{1/2}=O_p(K_N^{-r})+O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT})+o_p(N^{-1/2}), \tag{A.10}$$

$$\sup_{1\le t\le T}\{\int\eta_{N,jt}(x_j)^2dx_j\}^{1/2}=O_p(K_N^{-r})+O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT})+o_p(N^{-1/2}),$$
$$\sup_{1\le t\le T}\sup_{x_j\in[a,b]}|\eta_{N,jt}(x_j)|=O_p(K_N^{-r})+O_p(K_N^2N^{-3/4}\sqrt{\log NT})+o_p(N^{-1/2}). \tag{A.11}$$

Moreover, by Berntein's inequality and following the same procedure as the proof for Lemma **A.1**, we have $\sup_{1\le t\le T}||D_{Nt,1}||=O_p(\sqrt{K_N/N}\sqrt{\log K_NT})$. Hence,

$$\sup_{1\le t\le T}\sup_{x\in[a,b]^J}|1_{j+1}^{\intercal}\mathbb{B}(x)D_{Nt,1}|=O_p(\sqrt{\log K_NT}K_N/\sqrt{N}),$$
$$\sup_{1\le t\le T}\{N^{-1}\sum\nolimits_{i=1}^N(1_{j+1}^{\intercal}\mathbb{B}(X_i)D_{Nt,1})^2\}^{1/2}=O_p(\sqrt{\log K_NT}\sqrt{K_N/N}). \tag{A.12}$$

Therefore, by (A.9), (A.10), (A.11) and (A.12), we have

$$\sup_{1\le t\le T}N^{-1}\sum\nolimits_{i=1}^N\{\widetilde{h}_{jt}(X_{ji})-h_{jt}^0(X_{ji})\}^2=O_p((\log K_NT)K_N/N+N^{-2r}),$$

$$\sup_{1\le t\le T} \sup_{x_j\in[a,b]} |\widetilde{h}_{jt}(x_j) - h^0_{jt}(x_j)| = O_p(\sqrt{\log K_N T} K_N N^{-1/2} + K_N^{-r}). \quad (A.13)$$

Moreover, by Conditions (C3) and (C4), we have with probability approaching 1, as $N \to \infty$,

$$c_h \le N^{-1} \sum\nolimits_{i=1}^N (T^{-1} \sum\nolimits_{t=1}^T h^0_{jt}(X_{ji}))^2 \le C_h\ ,\ c_h \le N^{-1} \sum\nolimits_{i=1}^N (T^{-1} \sum\nolimits_{t=1}^T \widetilde{h}_{jt}(X_{ji}))^2 \le C_h. \quad (A.14)$$

Hence, this result together with (A.9) leads to that with probability approaching 1, as $N \to \infty$,

$$\begin{aligned}
&\left| 1/\sqrt{N^{-1} \sum\nolimits_{i=1}^N (T^{-1} \sum\nolimits_{t=1}^T \widetilde{h}_{jt}(X_{ji}))^2} - 1/\sqrt{N^{-1} \sum\nolimits_{i=1}^N (T^{-1} \sum\nolimits_{t=1}^T h^0_{jt}(X_{ji}))^2} \right| \\
&= \left| M_{NT} N^{-1} \sum\nolimits_{i=1}^N T^{-1} \sum\nolimits_{t=1}^T \{\widetilde{h}_{jt}(X_{ji}) - h^0_{jt}(X_{ji})\} T^{-1} \sum\nolimits_{t=1}^T \{\widetilde{h}_{jt}(X_{ji}) + h^0_{jt}(X_{ji})\} \right| \\
&= |2 M_{NT} N^{-1} \sum\nolimits_{i=1}^N T^{-1} \sum\nolimits_{t=1}^T \{\widetilde{h}_{jt}(X_{ji}) - h^0_{jt}(X_{ji})\} \{T^{-1} \sum\nolimits_{t=1}^T h^0_{jt}(X_{ji})\} \\
&+ M_{NT} N^{-1} \sum\nolimits_{i=1}^N T^{-1} \sum\nolimits_{t=1}^T \{\widetilde{h}_{jt}(X_{ji}) - h^0_{jt}(X_{ji})\}^2 | \\
&\le \left| 2 M_{NT} N^{-1} \sum\nolimits_{i=1}^N T^{-1} \sum\nolimits_{t=1}^T [1^\intercal_{j+1} \mathbb{B}(X_i) D_{Nt,1} \{T^{-1} \sum\nolimits_{t=1}^T h^0_{jt}(X_{ji})\} + \varrho_{it}] \right| \\
&+ 2 M_{NT} N^{-1} \sum\nolimits_{i=1}^N T^{-1} \sum\nolimits_{t=1}^T \{1^\intercal_{j+1} \mathbb{B}(X_i) D_{Nt,1}\}^2 + \eta_{N,jt}(X_{ji}) \\
&+ 2 M_{NT} N^{-1} \sum\nolimits_{i=1}^N T^{-1} \sum\nolimits_{t=1}^T \{\eta_{N,jt}(X_{ji})\}^2 \qquad (A.15)
\end{aligned}$$

for M_{NT} satisfying $M_{NT} \in (c', C')$ for some constants $0 < c' < C' < \infty$, where $\varrho_{it} = \eta_{N,jt}(X_{ji}) \{T^{-1} \sum_{t=1}^T h^0_{jt}(X_{ji})\}$. Moreover by (A.10), there exists a constant $C^* \in (0, \infty)$ such that

$$\begin{aligned}
|N^{-1} \sum\nolimits_{i=1}^N T^{-1} \sum\nolimits_{t=1}^T \varrho_{it}| &\le C^* \sup\nolimits_{1\le t\le T} N^{-1} \sum\nolimits_{i=1}^N |\eta_{N,jt}(X_{ji})| \\
&\le C^* \sup\nolimits_{1\le t\le T} [N^{-1} \sum\nolimits_{i=1}^N \{\eta_{N,jt}(X_{ji})\}^2]^{1/2} \\
&= O_p(K_N^{-r}) + O_p(K_N^{3/2} N^{-3/4} \sqrt{\log NT}) + o_p(N^{-1/2}), \qquad (A.16)
\end{aligned}$$

and

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

Define $\psi_{it} = \{\psi_{it,\ell}\}_{\ell=1}^{d(N)} = \Psi_{Nt}^{-1} Z_i(\tau - I(\varepsilon_{it} < 0))$. Then $E(\psi_{it,\ell}) = 0$. Moreover, $E||\psi_{it}||^2 \le c_1 K_N$for some constant $0 < c_1 < \infty$, and by Condition (C1), we have

$$\begin{aligned}|E(\psi_{it}^{\intercal}\psi_{js})| &\leq 2\{\phi(\sqrt{|i-j|^2+|t-s|^2})\}^{1/2}\sum\nolimits_{\ell=1}^{d(N)}\{E(\psi_{it,\ell})^2E(\psi_{js,\ell})^2\}^{1/2}\\ &\leq \{\phi(\sqrt{|i-j|^2+|t-s|^2})\}^{1/2}(E||\psi_{it}||^2+E||\psi_{js}||^2)\\ &\leq 2c_1K_N\{\phi(\sqrt{|i-j|^2+|t-s|^2})\}^{1/2}.\end{aligned}$$

Hence by Condition (C1), we have

$$\begin{aligned}&E||(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\psi_{it}||^2\\ &=(NT)^{-2}\sum\nolimits_{t,t'}\sum\nolimits_{i,i'}E(\psi_{it}^{\intercal}\psi_{i't'})\leq 2c_1K_N(NT)^{-2}\sum\nolimits_{t,t'}\sum\nolimits_{i,i'}\{\phi(\sqrt{|i-j|^2+|t-s|^2})\}^{1/2}\\ &\leq 2c_1K_1K_N(NT)^{-2}\sum\nolimits_{t,t'}\sum\nolimits_{i,i'}e^{-\lambda_1\sqrt{|i-i'|^2+|t-t'|^2}/2}\\ &\leq 2c_1(NT)^{-2}K_1K_N\sum\nolimits_{t,t'}\sum\nolimits_{i,i'}e^{-(\lambda_1/2)(|i-i'|+|t-t'|)}\\ &\leq 2c_1K_1K_N(NT)^{-2}(NT)(\sum\nolimits_{k=0}^{T}e^{-(\lambda_1/2)k})(\sum\nolimits_{k=0}^{N}e^{-(\lambda_1/2)k})\\ &\leq 2c_1K_1K_N(NT)^{-2}(NT)\{1-e^{-(\lambda_1/2)}\}^{-2}=2c_1K_1K_N\{1-e^{-(\lambda_1/2)}\}^{-2}(NT)^{-1}=O\{K_N(NT)^{-1}\}.\end{aligned}$$

Thus, by Markov's inequality, as $(N,T)\to\infty$,

$$||(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\psi_{it}||=O_p[\{K_N(NT)^{-1}\}^{1/2}]. \tag{A.18}$$

Moreover, by the definition of $D_{Nt,1}$ given in (A.5), we have $D_{Nt,1}=N^{-1}\sum_{i=1}^{N}\psi_{it}$. Therefore, as $(N,T)\to\infty$,

$$\begin{aligned}&|N^{-1}\sum\nolimits_{i=1}^{N}T^{-1}\sum\nolimits_{t=1}^{T}1_{j+1}^{\intercal}\mathbb{B}(X_i)D_{Nt,1}|=|N^{-1}\sum\nolimits_{i=1}^{N}1_{j+1}^{\intercal}\mathbb{B}(X_i)(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\psi_{it}|\\ &\leq ||(NT)^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\psi_{it}||[\lambda_{\max}\{N^{-1}\sum\nolimits_{i=1}^{N}B_j(X_{ji})B_j(X_{ji})^{\intercal}\}]^{1/2}=O_p[\{K_N(NT)^{-1}\}^{1/2}].\end{aligned}$$

By (A.12) and $\log(K_NT)K_NN^{-1/2}=o(1)$, we have

$$N^{-1}\sum\nolimits_{i=1}^{N}T^{-1}\sum\nolimits_{t=1}^{T}\{1_{j+1}^{\intercal}\mathbb{B}(X_i)D_{Nt,1}\}^2=\{O_p(\sqrt{\log K_NT}\sqrt{K_N/N})^2\}=o_p(N^{-1/2}).$$

Therefore, the above results together with (A.15), (A.16) and (A.17) lead to

$$\left|1/\sqrt{N^{-1}\sum_{i=1}^{N}(T^{-1}\sum_{t=1}^{T}\widetilde{h}_{jt}(X_{ji}))^2} - 1/\sqrt{N^{-1}\sum_{i=1}^{N}(T^{-1}\sum_{t=1}^{T}h_{jt}^0(X_{ji}))^2}\right|$$
$$= O_p[\{K_N(NT)^{-1}\}^{1/2}] + O_p(K_N^{-r}) + O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}).$$

Denote $\varpi_{NT} = \sqrt{N^{-1}\sum_{i=1}^{N}(T^{-1}\sum_{t=1}^{T}\widetilde{h}_{jt}(X_{ji}))^2}$ and $\varpi_{NT}^0 = \sqrt{N^{-1}\sum_{i=1}^{N}(T^{-1}\sum_{t=1}^{T}h_{jt}^0(X_{ji}))^2}$. Then

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\{\widetilde{h}_{jt}(x_j)/\varpi_{NT} - h_{jt}^0(x_j)/\varpi_{NT}^0\} \\
&= T^{-1}\sum_{t=1}^{T}\{\widetilde{h}_{jt}(x_j)/\varpi_{NT} - h_{jt}^0(x_j)/\varpi_{NT}\} + T^{-1}\sum_{t=1}^{T}\{h_{jt}^0(x_j)/\varpi_{NT} - h_{jt}^0(x_j)/\varpi_{NT}^0\} \\
&= T^{-1}\sum_{t=1}^{T}\{\widetilde{h}_{jt}(x_j) - h_{jt}^0(x_j)\}/\varpi_{NT} + T^{-1}\sum_{t=1}^{T}h_{jt}^0(x_j)\{1/\varpi_{NT} - 1/\varpi_{NT}^0\}.
\end{aligned}$$

By the above result and Condition (C3), we have

$$\begin{aligned}
&\sup_{x_j\in[a,b]}|T^{-1}\sum_{t=1}^{T}h_{jt}^0(x_j)\{1/\varpi_{NT} - 1/\varpi_{NT}^0\}| \\
&= O_p[\{K_N(NT)^{-1}\}^{1/2}] + O_p(K_N^{-r}) + O_p(K_N^{3/2}N^{-3/4}\sqrt{\log NT}) + o_p(N^{-1/2}).
\end{aligned}$$

Moreover, (A.9) leads to

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\{\widetilde{h}_{jt}(x_j) - h_{jt}^0(x_j)\}/\varpi_{NT} \\
&= T^{-1}\sum_{t=1}^{T}1_{j+1}^{\intercal}\mathbb{B}(x)D_{Nt,1}/\varpi_{NT} + T^{-1}\sum_{t=1}^{T}\eta_{N,jt}(x_j)/\varpi_{NT} = \Phi_{NTj,1}(x_j) + \Phi_{NTj,2}(x_j).
\end{aligned}$$

By (A.11) and (A.14), as $N\to\infty$,

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

By (A.14) and (A.18), as $(N,T)\to\infty$,

$$\begin{aligned}
\sup_{x_j\in[a,b]}|\Phi_{NTj,1}(x_j)| &\le c_h^{-1}\{||(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\psi_{it}||^2\sup_{x_j\in[a,b]}||B_j(x_j)||^2\}^{1/2} \\
&= O_p\{K_N(NT)^{-1/2}\}, \\
\left\{\int\Phi_{NTj,1}(x_j)^2dx_j\right\}^{1/2} &\le c_h^{-1}C_2\{||(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\psi_{it}||^2\}^{1/2} = O_p\{K_N^{1/2}(NT)^{-1/2}\}.
\end{aligned}$$

Hence, the results in Proposition 1 follow from the above results directly. ∎

9.2 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. We define the infeasible estimator $f_t^* = \{f_{ut}^*, (f_{jt}^*, 1 \le j \le J)^\intercal\}^\intercal$ as the minimizer of

$$\sum\nolimits_{i=1}^{N} \rho_\tau(y_{it} - f_{ut} - \sum\nolimits_{j=1}^{J} g_j^0(X_{ji}) f_{jt}). \tag{A.19}$$

Lemma 3. *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 (4.1).

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

$$f_t^* - f_t^0 = \Lambda_{Nt}^{-1}\{N^{-1} \sum\nolimits_{i=1}^{N} G_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\} + \upsilon_{Nt}, \tag{A.20}$$

and $||\upsilon_{Nt}|| = o_p(N^{-1/2})$ for every t, where $\Lambda_{Nt} = N^{-1} \sum_{i=1}^{N} p_{it}\,(0\,|X_i\,)\, G_i^0(X_i) G_i^0(X_i)^\intercal$. By Conditions (C2) 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 3 can be obtained directly by Condition (C6). ∎

Recall that the estimator $\widehat{f}_t^{[1]}$ given in (3.1) 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.19). Then we have the following result for $\widehat{f}_t^{[1]}$.

Lemma 4. *Let Conditions (C1)-(C5) hold. If, in addition, $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$, $K_N^{-1}(\log NT)(\log N)^4 = o(1)$ and $\phi_{NT} = o(1)$, where ϕ_{NT} is given in (4.2), then for a given t there is a stochastically bounded sequence $\delta_{N,jt}$ such that as $(N, T) \to \infty$,*

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

where $\delta_{N,t} = (\delta_{N,jt}, 0 \le j \le J)^\intercal$ and d_{NT} is given in (4.3).

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

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

so that f_t^* and $\widehat{f}_t^{[1]}$ are the minimizers of $L_{Nt}(f_t, g^0)$ and $L_{Nt}(f_t, \widehat{g}^{[0]})$, respectively, where $\widehat{g}^{[0]} = \{\widehat{g}_j^{[0]}(\cdot), 1 \leq j \leq J\}$ and $g^0 = \{g_j^0(\cdot), 1 \leq j \leq J\}$. 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 $g_j^0(x_j) = \widetilde{g}_j^0(x_j) + r_j(x_j)$,

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

By Proposition 1, 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 $||\boldsymbol{\lambda}_{j,NT} - \boldsymbol{\lambda}_j^0|| = O_p(d_{NT}) + o_p(N^{-1/2})$. Let d'_{NT} be a sequence satisfying $d'_{NT} = o(N^{-1/2})$ and let $d^*_{NT} = d_{NT} + d'_{NT}$. In the following, we will show that

$$\widetilde{f}_t - f_t^0 - d_{NT}\delta_{N,t} = \Lambda_{Nt}^{-1}\{N^{-1} \sum_{i=1}^{N} G_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\} + o_p(N^{-1/2}), \tag{A.21}$$

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C} d^*_{NT}$ for some constant $0 < \widetilde{C} < \infty$, where $\widetilde{f}_t$ is the minimizer of $L_{Nt}(f_t, g)$ and $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$. Hence the result in Lemma 4 follows from (A.20) and (A.21). We have $||\widetilde{f}_t - f_t^*|| = o_p(1)$, since

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

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 $\mathbf{X} = (X_1, \ldots, X_N)^\intercal$ and $G_i(X_i) = \{1, g_1(X_{1i}), \ldots, g_J(X_{Ji})\}^\intercal$. Let

$$\psi_\tau(\varepsilon) = \tau - I(\varepsilon < 0).$$

For $\boldsymbol{\lambda}_j \in R^{K_n}$ satisfying $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C} d_{NT}^*$ and f_t in a neighborhood of f_t^0, write

$$\begin{aligned} L_{Nt}(f_t, g) &= E\{L_{Nt}(f_t, g)|\mathbf{X}\} - (f_t - f_t^0)^\intercal \{W_{Nt,1} - E(W_{Nt,1}|\mathbf{X})\} \\ &+ W_{Nt,2}(f_t, g) - E(W_{Nt,2}(f_t, g)|\mathbf{X}), \end{aligned} \tag{A.22}$$

where $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$, and

$$W_{Nt,1} = N^{-1} \sum\nolimits_{i=1}^N G_i(X_i)\psi_\tau(y_{it} - f_t^{0\intercal} G_i(X_i)), \tag{A.23}$$

$$\begin{aligned} W_{Nt,2}(f_t, g) = N^{-1} \sum\nolimits_{i=1}^N &\{\rho_\tau(y_{it} - f_t^\intercal G_i(X_i)) - \rho_\tau(y_{it} - f_t^{0\intercal} G_i(X_i)) \\ &+ (f_t - f_t^0)^\intercal G_i(X_i)\psi_\tau(y_{it} - f_t^{0\intercal} G_i(X_i))\}. \end{aligned} \tag{A.24}$$

In Lemma **A.4** in the Supplemental Materials, we show that as $(N, T) \to \infty$,

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

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C} d_{NT}^*$ 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.22), we have with probability approaching 1,

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

In Lemma **A.5** in the Supplemental Materials, we show that

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

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C} d_{NT}$ and $||f_t - f_t^0|| \leq \varpi_N$. 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

$$\widetilde{f}_t - f_t^0 = \Lambda_{Nt}^{-1} W_{Nt,1} + o_p(N^{-1/2}). \tag{A.25}$$

In Lemma **A.6** in the Supplemental Materials, we show that for a given t there is a stochastically bounded sequence $\delta_{N,jt}$ such that as $(N, T) \to \infty$,

$$W_{Nt,1} = N^{-1} \sum\nolimits_{i=1}^N G_i^0(X_i)\psi_\tau(\varepsilon_{it}) + d_{NT}\delta_{N,t} + o_p(N^{-1/2}). \tag{A.26}$$

where $\delta_{N,t} = (\delta_{N,jt}, 0 \le j \le J)^\intercal$ and $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$, uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \le \widetilde{C} d_{NT}^*$. Hence, result (A.21) follows from (A.25) and (A.26) directly. Then the proof is complete. ∎

Let $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\intercal, \ldots, \boldsymbol{\lambda}_J^\intercal)^\intercal$. For given $\widehat{f}^{[1]}$, we obtain

$$\widehat{\boldsymbol{\lambda}}^{[1]} = (\boldsymbol{\lambda}_1^{[1]\intercal}, \ldots, \boldsymbol{\lambda}_J^{[1]\intercal})^\intercal = \arg\min_{\boldsymbol{\lambda}} \{(NT)^{-1} \sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T \rho_\tau(y_{it} - \widehat{f}_{ut}^{[1]} - \sum\nolimits_{j=1}^J B_j(X_{ji})^\intercal \boldsymbol{\lambda}_j \widehat{f}_{jt}^{[1]})\}.$$

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

$$\boldsymbol{\lambda}^* = (\boldsymbol{\lambda}_1^{*\intercal}, \ldots, \boldsymbol{\lambda}_J^{*\intercal})^\intercal = \arg\min_{\boldsymbol{\lambda}} \{(NT)^{-1} \sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T \rho_\tau(y_{it} - f_{ut}^0 - \sum\nolimits_{j=1}^J B_j(X_{ji})^\intercal \boldsymbol{\lambda}_j f_{jt}^0)\}.$$

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

$$\left[\int \{\widetilde{g}_j^*(x_j) - g_j^0(x_j)\}^2 dx_j\right]^{1/2} = O_p(K_N^{1/2}(NT)^{-1/2} + K_N^{-r}), \tag{A.27}$$

and

$$\int \{\widehat{g}_j^{[1]}(x_j)(x_j) - \widetilde{g}_j^*(x_j)\}^2 dx_j = O_p(d_{NT}^2) + o_p(N^{-1/2}). \tag{A.28}$$

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

$$\int \{\widehat{g}_j^{[1]}(x_j) - g_j^0(x_j)\}^2 dx_j = O_p(d_{NT}^2) + o_p(N^{-1/2}). \tag{A.29}$$

Proof. Denote $\widetilde{g}^0(x) = \{\widetilde{g}_j^0(x_j), 1 \le j \le J\}^\intercal$ and $g^*(x) = \{g_j^*(x_j), 1 \le j \le J\}^\intercal$. 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 $Q_{it}^0 = \{B_j(X_{ji})^\intercal f_{jt}^0, 1 \le j \le J\}^\intercal$,

$$\Psi_{NT} = (NT)^{-1} \sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T p_{it}\,(0\,|X_i\,)\, Q_{it}^0 Q_{it}^{0\intercal}, \tag{A.30}$$

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

$$U_{NT,1} = (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T Q^0_{it}(\tau - I(\varepsilon_{it} < 0)), \tag{A.31}$$
$$U_{NT,2} = (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T Q^0_{it}p_{it}(0|X_i)\left(\sum_{j=1}^J r^*_{j,it}\right).$$

By the same procedure as the proof of Lemma 2, for $K_N^4(\log(NT))^2(NT)^{-1} = o(1)$, we obtain the Bahadur representation for $\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^0$ as

$$\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^0 = \Psi^{-1}_{NT}(U_{N,1} + U_{N,2}) + R^*_{NT}, \tag{A.32}$$

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.32) and following the same reasoning as the proof for (A.13), 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.27) is proved. Define

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

Hence, $\widehat{\boldsymbol{\lambda}}^{[1]}$ and $\boldsymbol{\lambda}^*$ are the minimizers of $L^*_{NT}(\widehat{f}^{[1]},\boldsymbol{\lambda})$ and $L^*_{NT}(f^0,\boldsymbol{\lambda})$, respectively. In Lemma **A.7** in the Supplemental Materials, we show that

$$||\widehat{\boldsymbol{\lambda}}^{[1]} - \boldsymbol{\lambda}^0 - \Psi_{NT}^{-1} U_{N,1}|| = O_p(d_{NT}) + o_p(N^{-1/2}). \tag{A.33}$$

Hence, by (A.32), (A.33) and $||\Psi_{NT}^{-1} U_{N,2}|| = O(K_N^{-r})$, we have

$$||\widehat{\boldsymbol{\lambda}}^{[1]} - \boldsymbol{\lambda}^*|| = O_p(d_{NT}) + o_p(N^{-1/2}). \tag{A.34}$$

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

$$\{\sqrt{N^{-1}\sum\nolimits_{i=1}^N \widehat{g}_j^{*[1]}(X_{ji})^2}\}^{-1} - \{\sqrt{N^{-1}\sum\nolimits_{i=1}^N g_j^*(X_{ji})^2}\}^{-1} = O_p(d_{NT}) + o_p(N^{-1/2}).$$

Therefore, the result (A.28) follows from the above results directly. ∎

Proofs of Theorems 1 and 2. Based on (A.29) in Lemma 5, the result in Lemma 4 holds for $\widehat{f}_t^{[2]}$ with a different bounded sequence. Then the result (A.29) in Lemma 5 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 stops at the $(i+1)^{\text{th}}$ step for any finite number i, the results in Lemmas 4 and 5 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 3 and 4, directly, and Theorem 2 for $\widehat{g}_j$ is proved by using Lemma 5. ∎

9.3 Proofs of Theorem 3

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

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

and

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

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

$$D_{Nt,1} = (2Nh)^{-1} \sum\nolimits_{i=1}^{N} \{I(|\varepsilon_{it}| \leq h) - I(|\varepsilon_{it} - \widehat{d}_{it}(X_i)| \leq h)\} G_i^0(X_i) G_i^0(X_i)^\intercal,$$
$$D_{Nt,2} = (2Nh)^{-1} \sum\nolimits_{i=1}^{N} I(|\varepsilon_{it} - \widehat{d}_{it}(X_i)| \leq h)\{\widehat{G}_i(X_i)\widehat{G}_i(X_i)^\intercal - G_i^0(X_i) G_i^0(X_i)^\intercal\}.$$

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

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

where ϕ_{NT} is given in (4.2), 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,

$$\begin{aligned} E||\widehat{\Lambda}_{Nt} - \widetilde{\Lambda}_{Nt}|| &\leq c(2Nh)^{-1} \sum\nolimits_{i=1}^{N} E|\widehat{d}_{it}(X_i)| \times ||G_i^0(X_i) G_i^0(X_i)^\intercal|| \\ &\leq c(2Nh)^{-1} \sum\nolimits_{i=1}^{N} E\{\widehat{d}_{it}(X_i)\}^2 E||G_i^0(X_i) G_i^0(X_i)^\intercal||^2\}^{1/2} \\ &\leq c c_1^{1/2} (2Nh)^{-1} (\sqrt{K_N/(NT)} + K_N^{3/2} N^{-3/4} \sqrt{\log N} + K_N^{-r}) \times \\ &\sum\nolimits_{i=1}^{N} \{E||G_i^0(X_i) G_i^0(X_i)^\intercal||^2\}^{1/2}. \end{aligned}$$

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

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

for some constant $0 < C_a < \infty$. Hence, $||G_i^0(X_i) G_i^0(X_i)^\intercal|| \leq C_a$, and thus we have

$$\begin{aligned} E||\widehat{\Lambda}_{Nt} - \widetilde{\Lambda}_{Nt}|| &\leq c c_1^{1/2} (2Nh)^{-1} (\phi_{NT} + o(N^{-1/2})) \sum\nolimits_{i=1}^{N} C_a \\ &= 2^{-1} c c_1^{1/2} C_a h^{-1} (\phi_{NT} + o(N^{-1/2})) = o(1) \end{aligned}$$

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

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

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

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

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),

$$
\begin{aligned}
||E(\widetilde{\Lambda}_{Nt}-\Lambda_{Nt}^0)|| &= ||N^{-1}\sum_{i=1}^{N}E[\{(2h)^{-1}I(|\varepsilon_{it}|\leq h)-p_{it}\left(0\,|X_i\right)\}G_i^0(X_i)G_i^0(X_i)^{\intercal}]|| \\
&\leq c'h||N^{-1}\sum_{i=1}^{N}EQ_i^0(X_i)G_i^0(X_i)^{\intercal}|| = O(h)=o(1). \qquad \text{(A.35)}
\end{aligned}
$$

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

$$
\begin{aligned}
&\text{var}(\mathbf{a}^{\intercal}\widetilde{\Lambda}_{Nt}\mathbf{a}) \\
&=(2Nh)^{-2}\text{var}\left(\sum_{i=1}^{N}I(|\varepsilon_{it}|\leq h)\{a_0+\sum_{j=1}^{J}g_j^0(X_{ji})a_j\}^2\right) \\
&\leq (2Nh)^{-2}\sum_{i,i'}2\{\phi(|i-i'|)\}^{1/2}\times \\
&\left(E\left[I\left(|\varepsilon_{it}|\leq h\right)\{a_0+\sum_{j=1}^{J}g_j^0(X_{ji})a_j\}^4\right]\right)^{1/2}\left(E\left[I(|\varepsilon_{i't'}|\leq h)\{a_0+\sum_{j=1}^{J}g_j^0(X_{ji'})a_j\}^4\right]\right)^{1/2} \\
&\leq (J+1)^2\{a_0^2+C'^2a_j^2\}(2Nh)^{-2}(2C^*h)^2\sum_{i,i'}2\{\phi(|i-i'|)\}^{1/2} \\
&\leq (J+1)^2\{a_0^2+C'^2a_j^2\}N^{-2}2C^{*2}K_1\sum_{i,i'}e^{-(\lambda_1/2)(|i-i'|)} \\
&\leq (J+1)^2\{a_0^2+C'^2a_j^2\}2C^{*2}K_1N^{-1}\{1-e^{-(\lambda_1/2)}\}=O(N^{-1})=o(1). \qquad \text{(A.36)}
\end{aligned}
$$

By (A.35) and (A.36), we have $||\widetilde{\Lambda}_{Nt} - \Lambda_{Nt}^0|| = o_p(1)$. Hence,

$$||\widehat{\Lambda}_{Nt} - \Lambda_{Nt}^0|| \le ||\widehat{\Lambda}_{Nt} - \widetilde{\Lambda}_{Nt}|| + ||\widetilde{\Lambda}_{Nt} - \Lambda_{Nt}^0|| = o_p(1). \tag{A.37}$$

∎

9.4 Proofs of Theorem 4

Proof. Let $S_{[rN]t} = \sum_{i=1}^{[rN]} G_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))$, where $[a]$ denotes the largest integer no greater than a. Let $M = bN$. Define $\Lambda_{Nt}(r) = N^{-1}\sum_{i=1}^{[rN]} p_{it}\,(0\,|X_i\,)\,G_i^0(X_i)G_i^0(X_i)^\intercal$, $F_{Nt}(r) = N^{-1/2}S_{[rN]t}$, 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 $\widehat{w}_{Nt} = \frac{\tau(1-\tau)}{N}\sum_{i=1}^N \widehat{G}_i(X_i)\widehat{G}_i(X_i)^\intercal - N^{-1}\sum_{i=1}^N \widehat{v}_{it}\widehat{v}_{it}^\intercal$. Then

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

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 3-5, we have

$$\widehat{f}_t - f_t^0 = \Lambda_{Nt}^{-1}\{N^{-1}\sum_{i=1}^N G_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\} + o_p(N^{-1/2}), \tag{A.38}$$

$$\sup_{x_j\in\mathcal{X}_j} |\widehat{g}_j(x_j) - g_j^0(x_j)| = O_p(\phi_{NT}) + o_p(N^{-1/2}) = o_p(N^{-1/2}). \tag{A.39}$$

Let $r \in (0,1]$. Let $\widetilde{S}_{[rN]t} = \sum_{i=1}^{[rN]} G_i^0(X_i)(\tau - I(\widehat{\varepsilon}_{it}^0 < 0))$, where $\widehat{\varepsilon}_{it}^0 = y_{it} - \{\widehat{f}_{ut} + \sum_{j=1}^J g_j^0(X_{ji})\widehat{f}_{jt}\}$. By Lemma **A.6**, we have

$$||N^{-1/2}\widehat{S}_{[rN]t} - N^{-1/2}\widetilde{S}_{[rN]t}|| = o_p(1). \tag{A.40}$$

For any given $f_t \in R^{J+1}$, define $S_{[rN]t}(f_t) = \sum_{i=1}^{[rN]} G_i^0(X_i)(\tau - I(\varepsilon_{it}(f_t) < 0))$, 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 **A.8**, we have

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

Moreover,

$$
\begin{aligned}
&N^{-1/2}E[\{S_{[rN]t}(f_t) - S_{[rN]t}(f_t^0)\}|\mathbf{X}] \\
&= \sum\nolimits_{i=1}^{[rN]} G_i^0(X_i)E[(I(\varepsilon_{it}(f_t^0) < 0) - I(\varepsilon_{it}(f_t) < 0))|X_i, f_t],
\end{aligned} \tag{A.41}
$$

and thus by Taylor's expansion, we have

$$
\begin{aligned}
&||N^{-1/2}E[\{S_{[rN]t}(f_t) - S_{[rN]t}(f_t^0)\}|\mathbf{X}] \\
&-N^{-1/2}\sum\nolimits_{i=1}^{[rN]} p_{it}\left(0\,|X_i, f_t\right) G_i^0(X_i)G_i^0(X_i)^\intercal(f_t^0 - f_t)|| = o_p(1).
\end{aligned} \tag{A.42}
$$

Hence, by (A.40), (A.41) and (A.42), we have

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

This result, together with (A.38), implies

$$
N^{-1/2}\widehat{S}_{[rN]t} = \digamma_{Nt}(r) - \Lambda_{Nt}(r)\{\Lambda_{Nt}(1)\}^{-1}\digamma_{Nt}(1) + o_p(1). \tag{A.43}
$$

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.38) and (A.39), $\widehat{w}_{Nt} = o_p(1)$. 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

$$\widehat{\Omega}_{Nt,M=bN} = \sum_{i=1}^{N-1}\sum_{j=1}^{N-1}((K^*_{ij} - K^*_{i,j+1}) - (K^*_{i+1,j} - K^*_{i+1,j+1}))\frac{\widehat{S}_{it}}{\sqrt{N}}\frac{\widehat{S}_{jt}^{\intercal}}{\sqrt{N}} + o_p(1). \tag{A.44}$$

Moreover,

$$N^2((K^*_{ij} - K^*_{i,j+1}) - (K^*_{i+1,j} - K^*_{i+1,j+1})) = -D_{bN}\{(i-j)/N\}. \tag{A.45}$$

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 $\digamma_{Nt}(r) \xrightarrow{\mathcal{D}} W_{J+1}(r)\Upsilon_t^{\intercal}$. Thus,

$$(\Lambda_{Nt}(r), \digamma_{Nt}(r)^{\intercal}, D_{bN}(r)) \xrightarrow{\mathcal{D}} \left(r\Lambda_t^0, \Upsilon_t W_{J+1}(r)^{\intercal}, \frac{1}{b^2}K^{*\prime\prime}\left(\frac{r}{b}\right)\right). \tag{A.46}$$

Hence, by (A.43), (A.44), and (A.45), it follows that

$$\begin{aligned}\widehat{\Omega}_{Nt,M=bN} = &\int_0^1\int_0^1 -D_{bN}(r-s)[\digamma_{Nt}(r) - \Lambda_{Nt}(r)\{\Lambda_{Nt}(1)\}^{-1}\digamma_{Nt}(1)] \\ &\times [\digamma_{Nt}(s) - \Lambda_{Nt}(s)\{\Lambda_{Nt}(1)\}^{-1}\digamma_{Nt}(1)]^{\intercal} drds + o_p(1).\end{aligned} \tag{A.47}$$

By the continuous mapping theorem,

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

Then the proof is completed. ∎

9.5 Proofs of Theorems 5 and 6

Proof. By (A.38), $\widehat{f}_t - f_t^0 = N^{-1/2}\Lambda_{Nt}(1)^{-1}\digamma_{Nt}(1) + o_p(N^{-1/2})$. Then under H_0, we have

$$N^{1/2}(R\widehat{f}_t - r) = R\Lambda_{Nt}(1)^{-1}\digamma_{Nt}(1) + o_p(1). \tag{A.48}$$

It directly follows from (A.37), (A.46), (A.47) and (A.48) that

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

Since $R\Lambda_t^{0-1}\Upsilon_t W_{J+1}(1)$ is a $q\times 1$ vector of normal random variables with mean zero and variance $R\Lambda_t^{0-1}\Upsilon_t\Upsilon_t^\intercal\Lambda_t^{0-1}R^\intercal$, $R\Lambda_t^{0-1}\Upsilon_t W_{J+1}(1)$ can be written as $\Upsilon_t^* W_q(1)$, where $\Upsilon_t^*\Upsilon_t^{*\intercal} = R\Lambda_t^{0-1}\Upsilon_t\Upsilon_t^\intercal\Lambda_t^{0-1}R^\intercal$. Then replacing $R\Lambda_t^{0-1}\Upsilon_t W_{J+1}(1)$ 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

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

Thus by (A.46), we have

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

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

References

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

[2] Bai, J. and Li, K. (2012): "Statistical Analysis of Factor Models of High Dimension," *The Annals of Statistics*, 40,436-465.

[3] Bai, J. and Ng, S. (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70,191-221.

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

[5] 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.

[6] Bellman, R. E. (1961): "Adaptive Control Processes," Princeton University Press, Princeton, N.J..

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

[8] Bradley, B. O. and Taqqu, M. S. (2003): "Financial Risk and Heavy Tails," in book: Handbook of Heavy Tailed Distributions in Finance, 35-103.

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

[10] Cai, T. and Liu, W. (2011): "Adaptive Thresholding for Sparse Covariance Matrix Estimation," *Journal of the American Statistical Association*, 106, 672-684.

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

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

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

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

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

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

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

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

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

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

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

[22] Gorski, J., Pfeuffer, F. and Klamroth, K. (2007). "Biconvex sets and optimization with biconvex functions: a survey and extensions," Mathematical Methods of Operations Research, 66, 373-407.

[23] He, X. and Shi, P. (1996): "Bivariate tensor-product B-splines in a partly linear model," *Journal of Multivariate Analysis*, 58, 162-181.

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

[25] 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.

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

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

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

[29] Koenker, R., and Machado, J. (1999): "Goodness of Fite and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94, 1296-1310.

[30] Lam, C. and Yao, Q. (2012): "Factor Modeling for High-dimensional Time Series: Inference for the Number of Factors," *The Annals of Statistics*, 40, 694-726.

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

[32] Li, Q., Cheng, G., Fan, J., and Wang, Y. (2017): "Embracing Blessing of Dimensionality in Factor Models," *Journal of American Statistical Association*, forthcoming.

[33] Mammen, E. (1993): "Bootstrap and wild bootstrap for high dimensional linear models," *The Annals of Statistics*, 21, 255-285.

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

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

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

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

[38] 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.

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

[40] Xue and Yang. (2006): "Additive coefficient modeling via polynomial spline," *Statistica Sinica,* 16, 1423-1446.

Supplemental Materials for "Estimation and Inference in Semiparametric Quantile Factor Models"

Shujie Ma*
Department of Statistics, University of California Riverside
Oliver Linton
Department of Economics, University of Cambridge
Jiti Gao†
Department of Econometrics and Business Statistics, Monash University

In this supplement, we present Lemmas **A.1-A.3** which are used to prove Lemma 2 in Section 9.1. We also present Lemmas **A.4-A.6** which are used in the proofs of Lemmas 3 and 4, and Lemmas **A.7-A.8** which are used in the proofs of Lemma 5 in Section 9.2.

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

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

Proof. Let $B_N = \{\boldsymbol{\vartheta}_t : ||\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0|| \leq CK_N^{1/2}N^{-1/2}\}$. 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 $||(\boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}_t^0) - (\boldsymbol{\vartheta}_{t,v} - \boldsymbol{\vartheta}_t^0)|| \leq$

*The research of Ma is supported in part by the U.S. NSF grant DMS 1712558.
†The research of Gao is supported in part by the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421.

$C\{d(N)/N^{5/2}\}$ 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

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

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

$$\Delta_{tN,1} \le \max_{1\le v\le V} |\Gamma_{tN,1v}| + \max_{1\le v\le V} |\Gamma_{tN,2v}|, \qquad (A.2)$$

where

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

By Condition (C2), we have that there are some constants $0 < c', c'' < \infty$ such that

$$\sup_{1\le t\le T}\max_{1\le v\le V} |\Gamma_{tN,1v}| \le c'\gamma_N \max_{1\le i\le N} ||Z_i|| ||Z_i|| \le c''\{d(N)/N^{5/2}\}K_N = O(K_N^2 N^{-5/2}). \qquad (A.3)$$

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,

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

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

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

Hence

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

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

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

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

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

Hence, taking a large enough, one has

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

Then we have

$$\sup_{1\leq t\leq T}\max_{1\leq v\leq V}|\Gamma_{tN,2v}| = O_p\{K_N^{3/2}N^{-1}(\log NT)^3\}. \quad \text{(A.4)}$$

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 $|\widetilde{g}_{tN,i,\ell}(\boldsymbol{\vartheta}_{t,v})| \leq 4|Z_{i\ell}| \leq c_1 K_N^{1/2}$ for some constant $0 < c_1 < \infty$. Moreover,

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

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

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

Hence

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

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,

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

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

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

Hence,

$$P\left(\sup_{1\le t\le T}||N^{-1}\sum\nolimits_{i=1}^{N}\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_{t,v})-N^{-1}\sum\nolimits_{i=1}^{N}\widetilde{G}_{tN,i}(\boldsymbol{\vartheta}_{t}^{0})||\ge aK_N^{3/2}N^{-3/4}\sqrt{\log NT}\right)$$
$$\le d(N)T(NT)^{-c_3a^2K_N}.$$

By the union bound of probability again, we have

$$P\left(\sup_{1\le t\le T}|\Delta_{tN,2}|\ge aK_N^{3/2}N^{-3/4}\sqrt{\log NT}\right)\le(2N^2)^{d(N)}d(N)T(NT)^{-c_3a^2K_N}.$$

Hence, taking a large enough, one has

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

Then we have

$$\sup_{1\le t\le T}|\Delta_{tN,2}|=O_p\{K_N^{3/2}N^{-3/4}\sqrt{\log NT}\}. \tag{A.5}$$

Therefore, by (A.1), (A.2), (A.3), (A.4) and (A.5), we have

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

□

Lemma A.2. *Under Conditions (C1) and (C2), as* $N\to\infty$, $\sup_{1\le t\le T}||N^{-1}\sum_{i=1}^{N}G_{tN,i}(\widetilde{\boldsymbol{\vartheta}}_t)||=O_p(K_N^{3/2}N^{-1})$.

Lemma A.3. *Under Conditions (C2) and (C3), as* $N\to\infty$,

$$\Psi_{Nt}^{-1}G_{tN,i}^{*}(\boldsymbol{\vartheta}_t)=-(\boldsymbol{\vartheta}_t-\boldsymbol{\vartheta}_t^0)+N^{-1}\Psi_{Nt}^{-1}\sum\nolimits_{i=1}^{N}p_{it}\left(0\,|X_i\right)Z_ib_t(X_i)+R_{Nt}^{*},$$

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*.*

Proof. The proofs of Lemmas **A.2** and **A.3** follow the same procedure as in Lemmas A.4 and A.7 of Horowitz and Lee (2005) by using the results (A.1) and (A.3) which hold uniformly in $t = 1, ..., T$. □

Lemma A.4. *Under Conditions (C2) and (C3), as* $(N, T) \to \infty$*,*

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

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \le \widetilde{C} d^*_{NT}$ *and* $||f_t - f_t^0|| \le \varpi_N$*, where* $W_{Nt,1}$ *is defined in A.26 and* $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$*.*

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

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (I(u \le s) - I(u \le 0))ds,$$

we have

$$\begin{aligned}
&\rho_\tau(y_{it} - f_t^\intercal G_i(X_i)) - \rho_\tau(y_{it} - f_t^{0\intercal} G_i(X_i)) \\
&= -(f_t - f_t^0)^\intercal G_i(X_i)\psi_\tau(y_{it} - f_t^{0\intercal} G_i(X_i)) \\
&+ \int_0^{(f_t - f_t^0)^\intercal G_i(X_i)} \Big(I(y_{it} - f_t^{0\intercal} G_i(X_i) \le s) - I(y_{it} - f_t^{0\intercal} G_i(X_i) \le 0)\Big)\, ds. \qquad \text{(A.6)}
\end{aligned}$$

By Lipschitz continuity of $p_{it}(\varepsilon|X_i, f_t)$ given in Condition (C2) and boundedness of f^0_{jt} in Condition (C3), we have

$$\begin{aligned}
&F_{it}\{f_t^{0\intercal}(G_i(X_i) - G_i^0(X_i)) + s|X_i, f_t\} - F_{it}\{f_t^{0\intercal}(G_i(X_i) - G_i^0(X_i))|X_i\} \\
&= s p_{it}\{f_t^{0\intercal}(G_i(X_i) - G_i^0(X_i))|X_i\} + o(s),
\end{aligned}$$

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

$$
\begin{aligned}
& E\{L_{Nt}(f_t, g)|\mathbf{X}\} \\
& = -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}) + N^{-1} \sum_{i=1}^N \int_0^{(f_t - f_t^0)^\intercal G_i(X_i)} [F_{it}\{f_t^{0\intercal}(G_i(X_i) - G_i^0(X_i)) + s|X_i\} \\
& - F_{it}\{f_t^{0\intercal}(G_i(X_i) - G_i^0(X_i))|X_i\}]ds \\
& = -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}) + N^{-1} \sum_{i=1}^N \int_0^{(f_t - f_t^0)^\intercal G_i(X_i)} [sp_{it}\{f_t^{0\intercal}(G_i(X_i) - G_i^0(X_i))|X_i\}]ds \\
& + o\left[(f_t - f_t^0)^\intercal \{N^{-1} \sum_{i=1}^N G_i(X_i) G_i(X_i)^\intercal\}(f_t - f_t^0)\right] \\
& = -(f_t - f_t^0)^\intercal E(W_{Nt,1}|\mathbf{X}) + \frac{1}{2}(f_t - f_t^0)^\intercal \times \\
& \left[N^{-1} \sum_{i=1}^N p_{it}\{f_t^{0\intercal}(G_i(X_i) - G_i^0(X_i))|X_i\} G_i(X_i) G_i(X_i)^\intercal\right](f_t - f_t^0) \\
& + o\left[(f_t - f_t^0)^\intercal \{N^{-1} \sum_{i=1}^N G_i(X_i) G_i(X_i)^\intercal\}(f_t - f_t^0)\right]. \qquad (A.7)
\end{aligned}
$$

Since $\sup_{x_j \in [a,b]} |g_j(x_j) - g_j^0(x_j)| = o(1)$, then $\sup_{x \in \mathcal{X}} |f_t^{0\intercal}(G_i(x) - G_i^0(x))| = o(1)$. By similar reasoning to the proof for Theorem 2 in Lee and Robinson (2016), we have $N^{-1} \sum_{i=1}^N G_i(X_i) G_i(X_i)^\intercal = N^{-1} \sum_{i=1}^N E\{G_i(X_i) G_i(X_i)^\intercal\} + o_p(1)$. Hence, by these results, we have the result in Lemma **A.4**. □

Lemma A.5. *Under Conditions (C2) and (C3), we have, as* $(N, T) \to \infty$,

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

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C} d_{NT}^*$ *and* $||f_t - f_t^0|| \leq \varpi_N$, *where* $W_{Nt,2}(f_t, g)$ *is defined in (A.24) and* $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$.

Proof. By (A.6), we have

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

and thus

$$E(W_{Nt,2i}(f_t,g)|X_i)=\int_0^{(f_t-f_t^0)^\intercal G_i(X_i)}[F_{it}\{f_t^{0\intercal}(G_i(X_i)-G_i^0(X_i))+s|X_i\}$$
$$-F_{it}\{f_t^{0\intercal}(G_i(X_i)-G_i^0(X_i))|X_i\}]ds.$$

By following the same reasoning as the proof for (A.7), we have

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

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

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

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

$$E\{W_{Nt,2i}(f_t,g)\}^2$$
$$=E[E[\{\int_0^{(f_t-f_t^0)^\intercal G_i(X_i)}(I(y_{it}-f_t^{0\intercal}G_i(X_i)\le s)-I(y_{it}-f_t^{0\intercal}G_i(X_i)\le 0))ds\}^2|X_i]]$$
$$\le E[E[|I(y_{it}-f_t^{0\intercal}G_i(X_i)\le(f_t-f_t^0)^\intercal G_i(X_i))-I(y_{it}-f_t^{0\intercal}G_i(X_i)\le 0)|$$
$$\times\{(f_t-f_t^0)^\intercal G_i(X_i)\}^2|X_i]]$$
$$=E[E[|I(\varepsilon_{it}\le f_t{}^\intercal G_i(X_i)-f_t^{0\intercal}G_i(X_i)^0)-I(\varepsilon_{it}\le f_t^{0\intercal}(G_i(X_i)-G_i(X_i)^0)|$$
$$\times\{(f_t-f_t^0)^\intercal G_i(X_i)\}^2|X_i]]$$
$$\le C''E|(f_t-f_t^0)^\intercal G_i(X_i)|^3\le C'''E||f_t-f_t^0||^3$$

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

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

for some constant $0 < C'''' < \infty$. Following the same routine procedure as the proof in Lemma **A.1** by applying the Bernstein's inequality, we have

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

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

Lemma A.6. *Under Conditions (C1)-(C3), for a given t there is a stochastically bounded sequence $\delta_{N,jt}$ such that as $(N, T) \to \infty$,*

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

uniformly in $||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \leq \widetilde{C} d_{NT}^*$, where $W_{Nt,1}$ is defined in (A.23), $\delta_{N,t} = (\delta_{N,jt}, 0 \leq j \leq J)^\intercal$ and $g_j(x_j) = B_j(x_j)^\intercal \boldsymbol{\lambda}_j$.

Proof. Write

$$
W_{Nt,1} = W_{Nt,11} + W_{Nt,12} + W_{Nt,13}, \tag{A.8}
$$

where

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

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

$$\sup\nolimits_{||\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_j^0|| \le \widetilde{C} d_{NT}^*} |W_{Ntj,12}| = o_p(N^{-1/2}). \tag{A.9}$$

Moreover,

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

Let

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

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

$$E(W_{Ntj,13}|\mathbf{X}) = d_{NT}\delta_{N,jt} + o(N^{-1/2}). \tag{A.10}$$

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

Lemma **A.1**, we have

$$\sup_{||\boldsymbol{\lambda}_j-\boldsymbol{\lambda}_j^0||\leq\widetilde{C}d^*_{NT}}|W_{Ntj,13}-E(W_{Ntj,13}|\mathbf{X})| = o_p(N^{-1/2}). \tag{A.11}$$

Therefore, the result in Lemma **A.6** is proved by (A.8), (A.9), (A.10) and (A.11). □

Lemma A.7. *Let Conditions (C1)-(C4) hold. If, in addition,* $K_N^4N^{-1}=o(1)$, $K_N^{-r+2}(\log T)=o(1)$, $K_N^{-1}(\log NT)(\log N)^4=o(1)$ *and* $d_{NT}=o(1)$, *then we have as* $(N,T)\to\infty$,

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

where $U_{NT,1}$ *is defined in (A.31) and* Ψ_{NT} *is defined in (A.30).*

Proof. By Lemma 4 and (A.20), 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 $Q_{it}=\{B_j(X_{ji})^\intercal f_{jt},1\leq j\leq J\}^\intercal$. Let $f=(f_1^\intercal,\ldots,f_T^\intercal)^\intercal$ satisfy that $||f_t-f_t^0||\leq C_f(d_{NT}+N^{-1/2})$. Write

$$\begin{aligned}
&L^*_{NT}(f,\boldsymbol{\lambda})\\
&=E\{L^*_{NT}(f,\boldsymbol{\lambda})|\mathbf{X}\}-(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0)^\intercal\{V_{NT,1}(f)-E(V_{NT,1}(f)|\mathbf{X})\}\\
&+\ V_{NT,2}(f,\boldsymbol{\lambda})-E(V_{NT,2}(f,\boldsymbol{\lambda})|\mathbf{X}),
\end{aligned} \tag{A.12}$$

where

$$\begin{aligned}
V_{NT,1}(f)&=(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T Q_{it}\psi_\tau(y_{it}-f_{ut}-\boldsymbol{\lambda}^{0\intercal}Q_{it}), \qquad \text{(A.13)}\\
V_{NT,2}(f,\boldsymbol{\lambda})&=(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\{\rho_\tau(y_{it}-f_{ut}-\boldsymbol{\lambda}^\intercal Q_{it})-\rho_\tau(y_{it}-f_{ut}-\boldsymbol{\lambda}^{0\intercal}Q_{it})\\
&\quad+(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0)^\intercal Q_{it}\psi_\tau(y_{it}-f_{ut}-\boldsymbol{\lambda}^{0\intercal}Q_{it})\}.
\end{aligned}$$

By following the same reasoning as in the proofs of Lemmas **A.4** and **A.5**, we have

$$E\{L^*_{NT}(f,\boldsymbol{\lambda})|\mathbf{X}\}=-(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0)^\intercal E(V_{NT,1}(f)|\mathbf{X})+\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0)^\intercal\Psi_{NT}(\boldsymbol{\lambda}-\boldsymbol{\lambda}^0)+o_p(||\boldsymbol{\lambda}-\boldsymbol{\lambda}^0||^2), \tag{A.14}$$

$$V_{NT,2}(f,\boldsymbol{\lambda}) - E(V_{NT,2}(f,\boldsymbol{\lambda})|\mathbf{X}) = o_p(||\boldsymbol{\lambda}-\boldsymbol{\lambda}^0||^2 + (NT)^{-1}), \tag{A.15}$$

uniformly in $||f_t - f_t^0|| \le C_f(d_{NT} + N^{-1/2})$ and $||\boldsymbol{\lambda}-\boldsymbol{\lambda}^0|| \le \varsigma_{NT}$, where ς_{NT} is any sequence of positive numbers satisfying $\varsigma_{NT} = o(1)$.Thus, by (A.12), (A.14) and (A.15), 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 $||f_t - f_t^0|| \le C_f(d_{NT} + N^{-1/2})$ and $||\boldsymbol{\lambda}-\boldsymbol{\lambda}^0|| \le \varsigma_{NT}$. 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.3), as $(N,T)\to\infty$ with probability approaching 1, we have $||\Psi_{NT}^{-1}|| \le C'_\Psi$ for some constant $0 < C'_\Psi < \infty$. In Lemma **A.8**, 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 **A.7** follows from the above results, and thus the proof is completed. □

Lemma A.8. *Let Conditions (C1)-(C4) hold. If, in addition,* $K_N^4 N^{-1} = o(1)$, $K_N^{-r+2}(\log T) = o(1)$, $K_N^{-1}(\log NT)(\log N)^4 = o(1)$ *and* $d_{NT} = o(1)$, *then we have as* $(N,T)\to\infty$,

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

where $V_{NT,1}$ *and* $U_{NT,1}$ *are defined in (A.13) and (A.31), respectively.*

Proof. Write

$$V_{NT,1}(f) = V_{NT,11} + V_{NT,12}(f) + V_{NT,13}(f), \tag{A.16}$$

where

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

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,

$$\sup_{||f_t - f_t^0|| \le C_f(d_{NT}+N^{-1/2})} ||V_{NT,12}|| \le T^{-1}\sum_{t=1}^{T} ||N^{-1}\sum_{i=1}^{N} B(X_i)\psi_\tau(\varepsilon_{it})||$$
$$\times \sup_{||f_t - f_t^0|| \le C_f(d_{NT}+N^{-1/2})} ||f_t - f_t^0|| = O\{N^{-1/2}(d_{NT}+N^{-1/2})\} = o(N^{-1/2}+d_{NT}). \quad \text{(A.17)}$$

By following the same procedure as the proof for (A.36), 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 **A.1**, we have

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

Hence,

$$||V_{NT,13}(\widehat{f}^{[0]}) - E\{V_{NT,13}(\widehat{f}^{[0]})\}|| = o_p(d_{NT}). \quad \text{(A.18)}$$

Let

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

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

$$||E\{V_{NT,13}(f)|\mathbf{X}\}|| \le C||E[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} B_i(X_i)\{I(\varepsilon_{it} \le 0) - I(\varepsilon_{it} \le \kappa_{it}(f))\}|\mathbf{X}]||$$
$$\le C'||(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} B_i(X_i)\kappa_{it}(f)p_{it}(0|X_i, f_t)|| \quad \text{(A.19)}$$

uniformly in $||f_t - f_t^0|| \le C_f(d_{NT} + N^{-1/2})$. Moreover, by (A.20) and Lemma 4, we have

$$\Big\|(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)\kappa_{it}(\widehat{f}^{[0]})p_{it}(0|X_i) \\ +(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)p_{it}(0|X_i)\widetilde{g}^0(X_i)^{\intercal}[\Lambda_N^{-1}\{N^{-1}\sum\nolimits_{i=1}^{N} G_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\}]\Big\| \\ = O(d_{NT}) + o_p(N^{-1/2}). \tag{A.20}$$

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

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

we have

$$\Big\|(NT)^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} B_i(X_i)p_{it}(0|X_i)\widetilde{g}^0(X_i)^{\intercal}[\Lambda_N^{-1}\{N^{-1}\sum\nolimits_{i=1}^{N} G_i^0(X_i)(\tau - I(\varepsilon_{it} < 0))\}]\Big\| \\ = O_p\{(NT)^{-1/2}\}.$$

Therefore, by (A.19) and (A.20), we have with probability approaching 1,

$$||E\{V_{NT,13}(\widehat{f}^{[0]})|\mathbf{X}\}|| = O(d_{NT}) + o(N^{-1/2}). \tag{A.21}$$

By (A.18) and (A.21), we have

$$||V_{NT,13}(\widehat{f}^{[0]})|| = O_p(d_{NT}) + o_p(N^{-1/2}). \tag{A.22}$$

Therefore, the result in Lemma **A.8** follows from (A.16), (A.17), and (A.22) directly. □