Journal of Econometrics 170 (2012) 76–91

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Asymptotics for panel quantile regression models with individual effects[☆]

Kengo Kato[a,*], Antonio F. Galvao Jr.[b,c], Gabriel V. Montes-Rojas[d]

[a] Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
[b] Department of Economics, University of Iowa, W210 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242, United States
[c] Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201, United States
[d] Department of Economics, City University London, D306 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK

ARTICLE INFO

Article history:
Received 6 April 2010
Received in revised form
7 February 2012
Accepted 22 February 2012
Available online 7 April 2012

JEL classification:
C13
C21
C23

Keywords:
Asymptotics
Fixed effects
Panel data
Quantile regression

ABSTRACT

This paper studies panel quantile regression models with individual fixed effects. We formally establish sufficient conditions for consistency and asymptotic normality of the quantile regression estimator when the number of individuals, n, and the number of time periods, T, jointly go to infinity. The estimator is shown to be consistent under similar conditions to those found in the nonlinear panel data literature. Nevertheless, due to the non-smoothness of the objective function, we had to impose a more restrictive condition on T to prove asymptotic normality than that usually found in the literature. The finite sample performance of the estimator is evaluated by Monte Carlo simulations.

1. Introduction

Quantile regression (QR) for panel data has attracted considerable interest in both the theoretical and the applied literature. It allows us to explore a range of conditional quantiles, thereby exposing a variety of forms of conditional heterogeneity, and to control for unobserved individual effects. Controlling for individual heterogeneity via fixed effects, while exploring heterogeneous covariate effects within the QR framework, offers a more flexible approach to the analysis of panel data than that afforded by the classical Gaussian fixed and random effects estimation.

This paper focuses on the estimation of the common parameters in a QR model with individual effects. We refer the resulting estimator as the fixed effects quantile regression (FE-QR) estimator. Unfortunately, the FE-QR estimator is subject to the incidental parameter problem (Neyman and Scott (1948), Lancaster (2000), for a review) and will be inconsistent if the number of individuals n goes to infinity while the number of time periods T is fixed. It is important to note that, in contrast to mean regression, to our knowledge, there is no general transformation that can suitably eliminate the individual effects in the QR model. Therefore, given these difficulties, in the QR panel data literature, it is usual to allow T to increase to infinity to achieve asymptotically unbiased estimators. We follow this approach employing a large n, T asymptotics. In the nonlinear and quantile regression literature, the large panel data asymptotics is used in an attempt to cope with the incidental parameter problem.

The incidental parameter problem has been extensively studied in the recent nonlinear panel data literature. Among them, Hahn and Newey (2004) studied the maximum likelihood estimation of a general nonlinear panel data model with individual effects. They showed that the maximum likelihood estimator (MLE) has a limiting normal distribution with a bias in the mean when n and T grow at the same rate, and proposed several bias correction methods to the MLE. Note that since they assumed that likelihood functions are smooth, while the objective function of QR is not, their results are not directly applicable to the QR case.

Koenker (2004) introduced a novel approach for the estimation of a QR model for panel data. He argued that shrinking the

☆ We would like to express our appreciation to Roger Koenker, Hidehiko Ichimura, Kazuhiko Hayakawa, Ryo Okui, Stephane Bonhomme, Ivan Fernandez-Val, participants in the 16th Panel Data Conference, Amsterdam, and the 10th World Congress of the Econometric Society, Shanghai, for useful discussion and advice regarding this paper. We also would like to thank the editor, the associate editor, and three anonymous referees for their careful reading and comments to improve the manuscript. All the remaining errors are ours. Kato's research was supported by the Grant-in-Aid for Young Scientists (B) (22730179) from the JSPS.

* Corresponding author. Tel.: +81 82 424 7358.
E-mail addresses: kkato@hiroshima-u.ac.jp (K. Kato), antonio-galvao@uiowa.edu, agalvao@uwm.edu (A. F. Galvao Jr.), Gabriel.Montes-Rojas.1@city.ac.uk (G.V. Montes-Rojas).

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2012.02.007

individual parameters toward a common value improves the performance of the common parameters' estimates, and proposed a penalized estimation method where the individual parameters are subject to the ℓ_1 penalty. He also studied the asymptotic properties of the (unpenalized) FE-QR estimator, claiming its asymptotically normality when $n^a/T \to 0$ for some $a > 0$. We provide an alternative formal approach that offers a clearer understanding of the asymptotic properties of the FE-QR estimator and the related regularity conditions to establish these properties.

The goal of this paper is to study the asymptotic properties of the FE-QR estimator when n and T jointly go to infinity and formally establish sufficient conditions for consistency and asymptotic normality of the estimator. We show that the FE-QR estimator is consistent under similar conditions to those found in the nonlinear panel data literature. We are required to impose a more restrictive condition on T (i.e., $n^2(\log n)^3/T \to 0$) to prove asymptotic normality of the estimator than that found in the literature. This reflects the fact that the rate of the remainder term of the Bahadur representation of the FE-QR estimator is of order $(T/\log n)^{-3/4}$. The slower convergence rate of the remainder term is due to the non-smoothness of the scores. It is important to note that the growth condition on T for establishing $\sqrt{nT}$-consistency of the FE-QR estimator (or other fixed effects estimators in general) is determined so that it "kills" the remainder term. Thus, the rate of the remainder term is essential in the asymptotic analysis of the fixed effects estimation when n and T jointly go to infinity. The theoretical contribution of this paper is the rigorous study of the rate of the remainder term in the Bahadur representation of the FE-QR estimator, which we believe is far from trivial.

From a technical point of view, the proof of asymptotic normality of the FE-QR estimator is of independent interest. Because of the non-differentiability of the objective function, the stochastic expansion technique of Li et al. (2003) is no longer applicable to the asymptotic analysis of the FE-QR estimator. Instead, we adapt the Pakes and Pollard (1989) approach for proving asymptotic normality of the estimator. In addition, we make use of some inequalities from the empirical process literature (such as Talagrand's inequality) to establish the convergence rate of the remainder term in the Bahadur representation of the FE-QR estimator. These inequalities significantly simplify the proof. Our results are also extended to the case where temporal dependence is allowed.

From an applied perspective, however, the required rate condition for asymptotic normality might be seen as a negative result. The restrictive condition on T is not found in most of the panel data applications of interest. However, the paper highlights that special attention needs to be taken with respect to formal asymptotic study in the QR panel data (see the discussion in Section 3.2). In addition, it shows that small sample simulations are an important tool to study the estimator's performance.

We carried out Monte Carlo simulations to study the finite sample performance of the FE-QR estimator. The simulation study highlights some cases where the FE-QR estimator has large bias in panels with large n/T. In addition, the results show that, on the one hand, the estimated standard errors approximate the true ones very closely as the sample size increases, but on the other hand, the coverage probability of the asymptotic Gaussian confidence interval may be inaccurate when n/T is large. This is probably due to the fact that the variance of the FE-QR estimator decreases when nT increases while the bias decreases when T increases but is independent of n, so that the centering of the confidence interval will be severely distorted when n/T is large.

We now review the literature related to this paper. Lamarche (2010) studied Koenker's (2004) penalization method and discussed an optimal choice of the tuning parameter. Canay (2008) proposes a two-step estimator of the common parameters. The difference is that in his model, each individual effect is not allowed to change across quantiles. Graham et al. (2009) showed that when $T = 2$ and the explanatory variables are independent of the error term, the FE-QR estimator does not suffer from the incidental parameter problem. However, their argument does not apply to the general case. Rosen (2009) addressed a set identification problem of the common parameters when T is fixed. Chernozhukov et al. (2009) considered identification and estimation of the quantile structural function defined in Imbens and Newey (2009) of a nonseparable panel model with discrete explanatory variables. They studied bounds of the quantile structural function when T is fixed and the asymptotic behavior of the bounds when T goes to infinity.

This paper is organized as follows. In Section 2, we introduce a QR model with individual fixed effects and the FE-QR estimator we consider. In Section 3, we discuss the asymptotic properties of the FE-QR estimator. Proofs of the theorems in Section 3 are given in Appendix. In Section 4, we report a simulation study for assessing the finite sample performance of the FE-QR estimator. In Section 5 we extend the asymptotic results of Section 3 to the dynamic case where we allow for dependence across time. Finally, in Section 6 we present some discussion on the paper.

2. Quantile regression with individual effects

In this paper, we consider a QR model with individual effects

$$Q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_{i0}(\tau)) = \alpha_{i0}(\tau) + \boldsymbol{x}_{it}'\boldsymbol{\beta}_0(\tau) \tag{2.1}$$

where $\tau \in (0, 1)$ is a quantile index, y_{it} is a dependent variable, $\boldsymbol{x}_{it}$ is a p dimensional vector of explanatory variables, $\alpha_{i0}(\tau)$ is the i-th individual effect, and $Q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_{i0}(\tau))$ is the conditional τ-quantile of y_{it} given $(\boldsymbol{x}_{it}, \alpha_{i0}(\tau))$. In general, each $\alpha_{i0}(\tau)$ and $\boldsymbol{\beta}_0(\tau)$ can depend on τ, but we assume τ to be fixed throughout the paper and suppress such a dependence for notational simplicity, such that $\alpha_{i0}(\tau) = \alpha_{i0}$ and $\boldsymbol{\beta}_0(\tau) = \boldsymbol{\beta}_0$.[1] We make no parametric assumption on the relationship between α_{i0} and $\boldsymbol{x}_{it}$. Throughout the paper, the number of individuals is denoted by n and the number of time periods is denoted by $T = T_n$ that depends on n. In what follows, we omit the subscript n of T_n.

We consider the fixed effects estimation of $\boldsymbol{\beta}_0$, which is implemented by treating each individual effect also as a parameter to be estimated. Throughout the paper, as in Hahn and Newey (2004) and Fernandez-Val (2005), we treat α_{i0} as fixed by conditioning on them.[2] We consider the estimator $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ defined by

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) := \arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\rho_\tau(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta}), \tag{2.2}$$

where $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)'$ and $\rho_\tau(u) := \{\tau - I(u \le 0)\}u$ is the check function (Koenker and Bassett, 1978). Note that $\boldsymbol{\alpha}$ implicitly depends on n. We call $\hat{\boldsymbol{\beta}}$ the fixed effects quantile regression (FE-QR) estimator of $\boldsymbol{\beta}_0$. The optimization for solving (2.2) can be very large depending on n and T. However, as Koenker (2004) observed, in typical applications, the design matrix is very sparse. Standard sparse matrix storage schemes only require the

[1] In our model, the individual effects include the intercept term and the intercept term depends on the quantile. Thus, the individual effects depend on the quantile. Koenker (2004) used a different approach, where the individual specific intercepts are restricted to be the same across the quantiles. This procedure can be implemented using weighted QR, as proposed initially by Koenker (1984). It is important to note that both models are identical for our purposes of estimating a single fixed quantile.

[2] This treatment is similar to the interpretation of non-stochastic regressors.

space for the non-zero elements and their indexing locations. This considerably reduces the computational effort and memory requirements.

It is important to note that in the QR model, there is no general transformation that can suitably eliminate the individual effects. This intrinsic difficulty has been recognized by Abrevaya and Dahl (2008), among others, and was clarified by Koenker and Hallock (2000). They remarked that "Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects (p. 19)".

3. Asymptotic theory: static case

3.1. Main results

In this section, we investigate the asymptotic properties of the FE-QR estimator.

We first consider the consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. We say that $\hat{\boldsymbol{\alpha}}$ is weakly consistent if $\hat{\alpha}_i$ converges in probability to α_{i0} uniformly over $1 \leq i \leq n$, that is, $\max_{1\leq i\leq n} |\hat{\alpha}_i - \alpha_{i0}| \xrightarrow{p} 0$. We introduce some regularity conditions that ensure the consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$.

(A1) $\{(y_{it}, \boldsymbol{x}_{it}), t \geq 1\}$ is independent and identically distributed (i.i.d.) for each fixed i and independent across i.
(A2) $\sup_{i\geq 1} E[\|\boldsymbol{x}_{i1}\|^{2s}] < \infty$ for some real $s \geq 1$.

The distribution of $(y_{it}, \boldsymbol{x}_{it})$ is allowed to depend on i. Put $u_{it} := y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0$. Condition (A1) implies that $\{(u_{it}, \boldsymbol{x}_{it}), t \geq 1\}$ is i.i.d. for each fixed i and independent across i. Let $F_i(u|\boldsymbol{x})$ denote the conditional distribution function of u_{it} given $\boldsymbol{x}_{it} = \boldsymbol{x}$. We assume that $F_i(u|\boldsymbol{x})$ has density $f_i(u|\boldsymbol{x})$. Let $f_i(u)$ denote the marginal density of u_{it}.

(A3) For each $\delta > 0$,

$$\epsilon_\delta := \inf_{i\geq 1} \inf_{|\alpha|+\|\boldsymbol{\beta}\|_1=\delta} E\left[\int_0^{\alpha+\boldsymbol{x}_{i1}'\boldsymbol{\beta}} \{F_i(s|\boldsymbol{x}_{i1}) - \tau\} ds\right] > 0, \tag{3.1}$$

where $\|\cdot\|_1$ stands for the ℓ_1 norm.[3]

Condition (A1) is the same as Condition 1 (i) of Fernandez-Val (2005). Hahn and Newey (2004) also assume temporal and cross sectional independence. In condition (A1) we exclude temporal dependence to focus on the simplest case first and to highlight the difficulties arising from the FE-QR estimator. The present results are extended below (Section 5) to the dependent case under suitable mixing conditions as in Hahn and Kuersteiner (2004).[4] Condition (A2) corresponds to the moment condition of Fernandez-Val (2005, p. 12). Condition (A3) is an identification condition of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and corresponds to Condition 3 of Hahn and Newey (2004). In fact, it is sufficient for consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ that (3.1) is satisfied for any sufficiently small $\delta > 0$. Recall that $F_i(0|\boldsymbol{x}_{i1}) = \tau$. Under suitable integrability conditions, the expectation in (3.1) can be expanded as $(\alpha, \boldsymbol{\beta}')\Omega_i(\alpha, \boldsymbol{\beta}')' + o(\delta^2)$ for $|\alpha| + \|\boldsymbol{\beta}\|_1 = \delta$ uniformly over $i \geq 1$ as $\delta \to 0$, where $\Omega_i := E[f_i(0|\boldsymbol{x}_{i1})(1, \boldsymbol{x}_{i1}')(1, \boldsymbol{x}_{i1}')']$. If the minimum eigenvalue of Ω_i is bounded away from zero uniformly over $i \geq 1$, there exists a positive constant δ_0 such that for $0 < \delta \leq \delta_0$, (3.1) is satisfied. The following result states consistency. The proof is given in the Appendix.

Theorem 3.1. *Assume that $n/T^s \to 0$ as $n \to \infty$, where s is given in condition* (A2). *Then, under conditions* (A1)–(A3), $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *is weakly consistent.*

Theorem 3.1 is not covered by Hahn and Newey (2004) and Fernandez-Val (2005) because they assumed that the parameter spaces of α_{i0} and $\boldsymbol{\beta}_0$ are compact. In our problem, due to the convexity of the objective function, we can remove the compactness assumption of the parameter spaces. The condition on T in Theorem 3.1 is the same as that in Theorems 1–2 of Fernandez-Val (2005). If $\sup_{i\geq 1} \|\boldsymbol{x}_{i1}\| \leq M$ (a.s.) for some positive constant M, then the conclusion of the theorem holds when $\log n/T \to 0$ as $n \to \infty$. See Remark A.1 after the proof of Theorem 3.1 for details.

Next, we derive the limiting distribution of $\hat{\boldsymbol{\beta}}$. To this end, we consider another set of conditions.

(B1) There exists a constant M such that $\sup_{i\geq 1} \|\boldsymbol{x}_{i1}\| \leq M$ (a.s.).
(B2) (a) For each i, $f_i(u|\boldsymbol{x})$ is continuously differentiable with respect to u for each $\boldsymbol{x}$ and let $f_i^{(1)}(u|\boldsymbol{x}) := \partial f_i(u|\boldsymbol{x})/\partial u$; (b) there exist constants C_f and L_f such that $f_i(u|\boldsymbol{x}) \leq C_f$ and $|f_i^{(1)}(u|\boldsymbol{x})| \leq L_f$ uniformly over $(u, \boldsymbol{x})$ and $i \geq 1$; (c) $f_i(0)$ is bounded from below by some positive constant independent of i.
(B3) Put $\boldsymbol{\gamma}_i := E[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}]/f_i(0)$ and $\Gamma_n := n^{-1}\sum_{i=1}^n E[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}(\boldsymbol{x}_{i1}' - \boldsymbol{\gamma}_i')]$. (a) Γ_n is nonsingular for each n, and the limit $\Gamma := \lim_{n\to\infty} \Gamma_n$ exists and is nonsingular; (b) the limit $V := \lim_{n\to\infty} n^{-1}\sum_{i=1}^n E[(\boldsymbol{x}_{i1} - \boldsymbol{\gamma}_i)(\boldsymbol{x}_{i1} - \boldsymbol{\gamma}_i)']$ exists and is nonsingular.

Condition (B1) is assumed in Koenker (2004). This condition is used to ensure the "asymptotic" first order condition displayed in Eq. (A.7) in the proof of Theorem 3.2. Condition (B2) imposes some restrictions on the conditional densities and is standard in the QR literature (cf. Condition (ii) of Angrist et al. (2006, Theorem 3). Condition (B3) is concerned with the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. Condition (B3) (a) implies that the minimum eigenvalue of Γ_n is bounded away from zero uniformly over $n \geq 1$.

The term $\boldsymbol{\gamma}_i$ is the projection of $\boldsymbol{x}_{i1}$ onto the constant term 1 with respect to the norm $\|V\|^2 = E[f_i(0|\boldsymbol{x}_{i1})V^2]$ as $E[f_i(0|\boldsymbol{x}_{i1})(\boldsymbol{x}_{i1} - \boldsymbol{\gamma}_i)] = \mathbf{0}$, and has the same role as the mean $E[\boldsymbol{x}_{i1}]$ in the mean regression case.[5] More formally, the term $\boldsymbol{\gamma}_i$ comes from the fact that the lower $p \times (n+p)$ part of the inverse Hessian matrix of the expectation of the QR objective function in (2.2) evaluated at the truth is given by $\Gamma_n^{-1}[-\boldsymbol{\gamma}_1 \cdots - \boldsymbol{\gamma}_n\ I_p]$.

We now state the main theorem of the paper. The proof is given in the Appendix.

Theorem 3.2. *Assume conditions* (A1), (A3) *and* (B1)–(B3). *If $\log n/T \to 0$ as $n \to \infty$ but T grows at most polynomially in n, then $\hat{\boldsymbol{\beta}}$ admits the expansion*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) = \Gamma_n^{-1}\left[\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \{\tau - I(u_{it} \leq 0)\}(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)\right] + O_p\{(T/\log n)^{-3/4}\}. \tag{3.2}$$

---

[3] There is no significant role in the ℓ_1 norm, as any norm on a fixed dimensional Euclidean space is equivalent. The ℓ_1 norm is used just to avoid the notation like $\|(\alpha_i - \alpha_{i0}, \boldsymbol{\beta}' - \boldsymbol{\beta}_0')'\|$.

[4] The independence assumption is used mainly to apply some standard stochastic inequalities; our results are extended below to the dependent case by replacing these stochastic inequalities by those that hold under suitable dependence conditions. We shall mention that the condition on T for the mean-zero asymptotic normality, which is given in Theorem 3.2, is not weakened when the observations are temporally dependent.

[5] The norm $\|V\|^2 = E[f_i(0|\boldsymbol{x}_{i1})V^2]$ is a Fisher-like norm to the QR objective function, as $E[f_i(0|\boldsymbol{x}_{i1})V^2] = d^2E[\rho_\tau(u_{i1} - tV)]/dt^2$.

If moreover $n^2(\log n)^3/T \to 0$, *then we have*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\{\mathbf{0}, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}\}.$$

The restriction that T grows at most polynomially in n is only to simplify the exposition, as it ensures $\log T = O(\log n)$. We shall stress that the Bahadur representation (3.2) is valid without the condition that $n^2(\log n)^3/T \to 0$. This condition is used only to "kill" the remainder term (the second term on the right side of (3.2)). Some other specific comments are listed in the next subsection.

We now turn to estimate the asymptotic covariance matrix. The estimation of Γ and V depends on the conditional densities, and therefore, they are not directly estimated by their sample analogues because the conditional densities are unknown. We consider the kernel estimation of the matrices Γ and V. Let $K: \mathbb{R} \to \mathbb{R}$ denote a kernel function (probability density function). Let $\{h_n\}$ denote a sequence of positive numbers (bandwidths) such that $h_n \to 0$ as $n \to \infty$. We use the notation $K_{h_n}(u) = h_n^{-1}K(u/h_n)$. Let $\hat{u}_{it} = y_{it} - \hat{\alpha}_i - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}$, which can be viewed as an "estimator" of u_{it}. It is seen that Γ and V can be estimated by

$$\hat{\Gamma} := \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{x}_{it}(\boldsymbol{x}_{it} - \hat{\boldsymbol{\gamma}}_i)',$$

$$\hat{V} := \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \hat{\boldsymbol{\gamma}}_i)(\boldsymbol{x}_{it} - \hat{\boldsymbol{\gamma}}_i)',$$

where

$$\hat{f}_i := \frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it}), \qquad \hat{\boldsymbol{\gamma}}_i := \frac{1}{\hat{f}_i T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{x}_{it}.$$

To guarantee the consistency of $\hat{\Gamma}$ and $\hat{V}$, we assume the following.

(C1) The kernel K is continuous, bounded and of bounded variation on $\mathbb{R}$.

(C2) $h_n \to 0$ and $\log n/(Th_n) \to 0$ as $n \to \infty$.

Condition (C1) is an assumption we only make on the kernel. Most standard kernels such as Gaussian and Epanechnikov kernels satisfy condition (C1). Although the uniform kernel does not satisfy condition (C1) as it is not continuous, the continuity of the kernel is used only to ensure that the class of functions $\{(u, \boldsymbol{x}) \mapsto K((u-\alpha-\boldsymbol{x}'\boldsymbol{\beta})/h_n) : (\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}\}$ is pointwise measurable, and it is verified that the uniform kernel also ensures this property.[6] Condition (C2) is a restriction on the bandwidth h_n. The bandwidth h_n needs to be slightly slower than T^{-1}.

Proposition 3.1. *Assume conditions* (A1), (A3), (B1)–(B3) *and* (C1)–(C2). *If* T *grows at most polynomially in* n, *we have* $\hat{\Gamma} \xrightarrow{p} \Gamma$ *and* $\hat{V} \xrightarrow{p} V$.

We shall mention that the consistency of $\hat{\Gamma}$ and $\hat{V}$ only requires the consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$, which is guaranteed by conditions (A1), (A3), (B1) and (C2) (observe that condition (C2) implies that $\log n/T \to 0$). It is now straightforward to see that the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$, $\tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}$, is consistently estimated by $\tau(1-\tau)\hat{\Gamma}^{-1}\hat{V}\hat{\Gamma}^{-1}$.

3.2. Discussion on Theorem 3.2

In this subsection, we give some discussion on Theorem 3.2.

1. *Relation to* Hahn and Newey (2004): Eqs. (10) and (17) in Hahn and Newey (2004) show that the MLE of the common parameters for smooth likelihood functions admits the representation

$$\hat{\theta} - \theta_0 = \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right) + \frac{1}{2T}\theta^{\epsilon\epsilon}(0) + \frac{1}{6T^{3/2}}\theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}), \tag{3.3}$$

where $\hat{\theta}, \theta_0, \mathcal{I}_i, U_{it}, \theta^{\epsilon\epsilon}(\cdot)$ and $\theta^{\epsilon\epsilon\epsilon}(\cdot)$ are defined in Hahn and Newey (2004) and $\tilde{\epsilon}$ is in $[0, T^{-1/2}]$. Under suitable regularity conditions, $\theta^{\epsilon\epsilon}(0)$ is $O_p(1)$ and $\theta^{\epsilon\epsilon\epsilon}(\epsilon)$ is $O_p(1)$ uniformly over $\epsilon \in [0, T^{-1/2}]$, which implies that the last two terms on the right side of Eq. (3.3) are $O_p(T^{-1})$ and $O_p(T^{-3/2})$, respectively.[7]

The difference from their result is that the rate of the remainder term of the FE-QR estimator (the second term on the right side of (3.2)) is roughly $T^{-3/4}$, which is significantly slower than T^{-1}. Hahn and Newey (2004) assumed that the scores are sufficiently smooth with respect to the parameters. On the other hand, the scores for problem (2.2), which are formally defined in Appendix, are not differentiable (in fact they consist of indicator functions). This means that, in contrast to estimators with smooth objective functions that have been studied in the literature such as Li et al. (2003), Hahn and Newey (2004) and Fernandez-Val (2005), the Taylor-series methods of asymptotic distribution theory do not apply to the FE-QR estimator, which greatly complicates the analysis of its asymptotic distributional properties. The difficulty is partly explained by the fact that, as Hahn and Newey (2004) observed, the first order asymptotic behavior of the (smooth) MLE of the common parameters can be affected by the second order behavior of the estimators of the individual parameters, while the second order behavior of QR estimators is non-standard and rather complicated (Arcones, 1998; Knight, 1998). In particular, for cross-sectional models, the second order of the QR estimator is $n^{-3/4}$ and not n^{-1} when the sample size is n. We shall mention that our proof strategy leads to the standard condition (i.e., $n/T \to 0$) up to the log term for the mean-zero asymptotic normality when the scores are smooth (see the remark after the proof of Theorem 3.2 for the technical reason why the slower rate appears).

However, it should be pointed out that although the above rate of the remainder term is the best one (up to the log term) that we could achieve, there might be a room for improvement on the rate, which means that our condition for the asymptotic normality is only a sufficient one. It is an open question whether the mean-zero asymptotic normality holds under the standard assumption that $n/T \to 0$.

2. *Relation to* Koenker (2004): Koenker (2004) claimed asymptotic normality of the FE-QR estimator under similar conditions to ours except that he assumed that $n^a/T \to 0$ for some $a > 0$. We believe that our proof of asymptotic normality offers a clearer understanding of the asymptotic properties of the FE-QR estimator than that in his Theorem 1. Actually, in his proof, a formal proof for $\sqrt{nT}$-consistency of $\hat{\boldsymbol{\beta}}$ is not offered, and a justification for the second expression of R_{mn} in p. 82 when n and m (in his notation) jointly go to infinity is not presented.

3. *Relation to* He and Shao (2000): He and Shao (2000) studied a general M-estimation with diverging number of parameters that

---

[6] See Appendix B for the definition of the pointwise measurability.

[7] In fact, Hahn and Newey (2004) showed that $\theta^{\epsilon\epsilon}(0)$ converges in probability to some constant vector, which will contribute to the bias in the asymptotic distribution when n and T grow at the same rate.

Table 1
Bias of $\hat{\beta}(\tau)$. Location shift model.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100				
0.25	25	0.003	0.000	0.000	0.000	0.004	0.003	0.001	0.000	−0.002	−0.002	−0.001	−0.001
		[0.056]	[0.037]	[0.016]	[0.011]	[0.075]	[0.048]	[0.021]	[0.015]	[0.122]	[0.075]	[0.031]	[0.023]
	50	0.000	0.000	0.000	0.000	0.003	0.002	0.000	0.000	−0.004	−0.001	−0.001	−0.000
		[0.040]	[0.026]	[0.011]	[0.008]	[0.051]	[0.034]	[0.015]	[0.011]	[0.080]	[0.050]	[0.022]	[0.016]
	100	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	−0.001	0.000	−0.000
		[0.028]	[0.018]	[0.008]	[0.006]	[0.035]	[0.023]	[0.010]	[0.007]	[0.055]	[0.035]	[0.016]	[0.011]
	200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	0.000	−0.000	0.000
		[0.019]	[0.013]	[0.006]	[0.004]	[0.024]	[0.017]	[0.007]	[0.005]	[0.037]	[0.025]	[0.011]	[0.008]
0.50	25	0.000	0.000	0.000	0.000	0.006	0.002	0.001	0.000	0.001	0.000	−0.000	0.000
		[0.051]	[0.035]	[0.015]	[0.010]	[0.099]	[0.073]	[0.032]	[0.022]	[0.087]	[0.053]	[0.019]	[0.013]
	50	0.000	0.000	0.000	0.000	0.004	0.002	0.000	0.000	−0.001	0.001	−0.000	0.000
		[0.036]	[0.025]	[0.011]	[0.007]	[0.068]	[0.051]	[0.022]	[0.015]	[0.059]	[0.036]	[0.014]	[0.009]
	100	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	−0.000	0.000	−0.000
		[0.025]	[0.017]	[0.007]	[0.005]	[0.047]	[0.035]	[0.015]	[0.011]	[0.040]	[0.025]	[0.009]	[0.007]
	200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.000	0.000	0.000	0.000
		[0.017]	[0.012]	[0.005]	[0.004]	[0.034]	[0.025]	[0.011]	[0.008]	[0.028]	[0.018]	[0.007]	[0.005]
0.75	25	−0.001	−0.001	0.000	0.000	0.005	0.004	0.001	0.000	0.002	0.003	0.001	0.000
		[0.056]	[0.037]	[0.016]	[0.011]	[0.154]	[0.105]	[0.048]	[0.034]	[0.121]	[0.074]	[0.031]	[0.022]
	50	−0.001	0.000	0.000	0.000	0.003	0.001	0.000	0.000	−0.001	0.002	0.000	0.000
		[0.039]	[0.026]	[0.011]	[0.008]	[0.106]	[0.073]	[0.034]	[0.024]	[0.080]	[0.051]	[0.022]	[0.016]
	100	−0.001	0.000	0.000	0.000	−0.001	−0.001	0.000	0.000	0.001	0.000	0.000	0.000
		[0.028]	[0.018]	[0.008]	[0.006]	[0.073]	[0.051]	[0.024]	[0.017]	[0.054]	[0.035]	[0.015]	[0.011]
	200	0.000	0.000	0.000	0.000	−0.001	−0.001	0.000	−0.001	−0.000	0.000	0.000	0.000
		[0.020]	[0.013]	[0.005]	[0.004]	[0.052]	[0.037]	[0.017]	[0.012]	[0.037]	[0.025]	[0.011]	[0.008]

Notes: Monte Carlo experiments based on 5000 repetitions. Standard deviations in brackets.

allows for non-smooth objective functions. It is interesting to note that their Corollary 3.2 shows that the smoothness of scores is crucial for the growth condition of the number of parameters in asymptotic distribution theory of M-estimators. However, it should be pointed out that our Theorem 3.2 is not derived from their result because of the specific nature of the panel data problem. The formal problem to apply their result is that the convergence rate of $\hat{\alpha}_i$ is different from that of $\hat{\boldsymbol{\beta}}$. To avoid this, make a reparametrization $\boldsymbol{\theta} = (n^{-1/2}\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ and put $\boldsymbol{z}_{it} := (n^{1/2}\boldsymbol{e}_i', \boldsymbol{x}_{it}')'$, where $\boldsymbol{e}_i$ is the i-th unit vector in $\mathbb{R}^n$. Then, the current problem is under the framework of He and Shao (2000) with $x_i = (y_{it}, \boldsymbol{z}_{it}), m = (n+p), p = (n+p), n = nT, \theta = \boldsymbol{\theta}$ and $\psi(x_i, \theta) = \{\tau - I(y_{it} \le \boldsymbol{z}_{it}'\boldsymbol{\theta})\}\boldsymbol{z}_{it}$.[8] Although conditions (C0)–(C3) may be achieved in this case, it is difficult to obtain a tight bound of $A(n, m)$ in conditions (C4) and (C5) of their paper. If we use the same reasoning as in Lemma 2.1 of He and Shao (2000), $A(n, m)$ is bounded by a constant times $n^{3/2}T^{1/2}$ (in our notation), but if we use this bound, the condition on T implied by Theorem 2.2 of He and Shao (2000) will be such that $n^3(\log n)^2/T \to 0$.

4. *On the proof of* Theorem 3.2: The proof of Theorem 3.2 is of independent interest. The proof proceeds as follows. It is based on the method of Pakes and Pollard (1989), but requires some extra efforts. The first step is to obtain certain representations of $\hat{\alpha}_i - \alpha_{i0}$ by expanding the first n elements of the scores. Plugging them into the expansion of the last p elements of the scores, we obtain a representation of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ (see (A.5)). The remaining task is to evaluate the remainder terms in the representation of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$, which corresponds to establishing the stochastic equicontinuity condition in Pakes and Pollard (1989). However, since the number of parameters goes to infinity as $n \to \infty$, the "standard" empirical process argument such as that displayed in their paper will not suffice to show this. In order to establish the convergence rate of the remainder terms, we make use of some empirical process techniques such as celebrated Talagrand's (1996) inequality, which significantly simplify the proof.

[8] The left sides correspond to the notation of He and Shao (2000) and the right sides correspond to our notation.

4. Monte Carlo

We investigate the finite sample performance of the FE-QR estimator. Two simple versions of model (2.1) are considered in the simulation study:

1. Location shift model: $y_{it} = \eta_i + x_{it} + \epsilon_{it}$;
2. Location-scale shift model: $y_{it} = \eta_i + x_{it} + (1 + \gamma x_{it})\epsilon_{it}$,

where $x_{it} = 0.3\eta_i + z_{it}$, $z_{it} \sim$ i.i.d. χ^2_3, $\eta_i \sim$ i.i.d. $U[0, 1]$ and $\epsilon_{it} \sim$ i.i.d. F with $F = N(0, 1)$, χ^2_3 or Cauchy. In the location shift model, $\alpha_{i0} = \alpha_{i0}(\tau) = \eta_i + F^{-1}(\tau)$ and $\beta_0(\tau) = 1$, while in the location-scale shift model, $\alpha_{i0} = \alpha_{i0}(\tau) = \eta_i + F^{-1}(\tau)$ and $\beta_0 = \beta_0(\tau) = 1 + \gamma F^{-1}(\tau)$. We consider cases where $n \in \{25, 50, 100, 200\}$, $T \in \{5, 10, 50, 100\}$ and $\tau \in \{0.25, 0.50, 0.75\}$. For the location-scale shift model we use $\gamma \in \{0.5, 1\}$.

Tables 1, 4 and 7 report the bias and the standard deviation of the FE-QR estimator. Tables 2, 5 and 8 report the average of the estimated standard error (together with its standard deviation) described in Proposition 3.1. Finally, the empirical coverage probability of the asymptotic Gaussian confidence interval at the 95% nominal level is constructed using this estimated standard error (Tables 3, 6 and 9). The empirical coverage probability is also computed. The number of Monte Carlo repetitions is 5000 in all cases.

4.1. Bias

The performance of the FE-QR estimator is evaluated first by its bias. Tables 1, 4 and 7 report the results for the location shift and location-scale shift ($\gamma = 0.5, 1$) models, respectively. For the median, the results are in line with those of Koenker (2004), where in both models the FE-QR estimator has small bias and standard deviation in small samples. However, there are noticeable differences for the first and third quartiles. In the location shift model, the bias is small in every case and the standard errors decrease monotonically as either n or T increases. In the location-scale shift model, however, both bias and standard errors are large for small T. In particular, the bias is considerable in the Cauchy and

Table 2
Estimated standard errors of $\hat{\beta}(\tau)$. Location shift model.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100				
0.25	25	0.061	0.042	0.016	0.011	0.122	0.078	0.026	0.017	0.132	0.085	0.030	0.021
		[0.013]	[0.006]	[0.002]	[0.001]	[0.027]	[0.012]	[0.002]	[0.001]	[0.040]	[0.018]	[0.004]	[0.002]
	50	0.040	0.028	0.012	0.008	0.076	0.049	0.017	0.011	0.083	0.056	0.021	0.015
		[0.007]	[0.004]	[0.001]	[0.000]	[0.012]	[0.006]	[0.001]	[0.001]	[0.018]	[0.009]	[0.002]	[0.001]
	100	0.028	0.019	0.008	0.006	0.049	0.032	0.011	0.008	0.055	0.038	0.015	0.011
		[0.004]	[0.002]	[0.000]	[0.000]	[0.006]	[0.003]	[0.001]	[0.000]	[0.009]	[0.005]	[0.001]	[0.001]
	200	0.019	0.014	0.006	0.004	0.033	0.021	0.008	0.005	0.037	0.026	0.011	0.008
		[0.002]	[0.001]	[0.000]	[0.000]	[0.003]	[0.001]	[0.000]	[0.000]	[0.005]	[0.003]	[0.001]	[0.000]
0.50	25	0.063	0.042	0.016	0.011	0.129	0.084	0.032	0.022	0.127	0.075	0.024	0.016
		[0.012]	[0.006]	[0.001]	[0.001]	[0.027]	[0.012]	[0.002]	[0.001]	[0.032]	[0.013]	[0.002]	[0.001]
	50	0.041	0.027	0.011	0.008	0.083	0.054	0.022	0.015	0.080	0.047	0.016	0.011
		[0.006]	[0.003]	[0.001]	[0.000]	[0.013]	[0.006]	[0.001]	[0.001]	[0.015]	[0.006]	[0.001]	[0.001]
	100	0.028	0.019	0.008	0.005	0.055	0.036	0.015	0.011	0.052	0.031	0.011	0.007
		[0.003]	[0.002]	[0.000]	[0.000]	[0.007]	[0.003]	[0.001]	[0.000]	[0.007]	[0.003]	[0.001]	[0.000]
	200	0.019	0.013	0.005	0.004	0.037	0.025	0.011	0.008	0.035	0.021	0.007	0.005
		[0.002]	[0.001]	[0.000]	[0.000]	[0.004]	[0.002]	[0.000]	[0.000]	[0.004]	[0.002]	[0.000]	[0.000]
0.75	25	0.061	0.042	0.017	0.011	0.136	0.100	0.046	0.033	0.132	0.085	0.030	0.021
		[0.013]	[0.006]	[0.002]	[0.001]	[0.034]	[0.019]	[0.006]	[0.003]	[0.038]	[0.018]	[0.004]	[0.002]
	50	0.040	0.028	0.012	0.008	0.092	0.070	0.033	0.023	0.084	0.056	0.021	0.015
		[0.007]	[0.004]	[0.001]	[0.001]	[0.018]	[0.011]	[0.003]	[0.002]	[0.018]	[0.009]	[0.002]	[0.001]
	100	0.028	0.020	0.008	0.006	0.064	0.050	0.024	0.017	0.055	0.038	0.015	0.011
		[0.004]	[0.002]	[0.001]	[0.000]	[0.010]	[0.006]	[0.002]	[0.001]	[0.009]	[0.005]	[0.001]	[0.001]
	200	0.019	0.014	0.006	0.004	0.045	0.036	0.017	0.012	0.037	0.026	0.011	0.008
		[0.002]	[0.001]	[0.000]	[0.000]	[0.006]	[0.004]	[0.001]	[0.001]	[0.005]	[0.003]	[0.001]	[0.000]

Notes: Monte Carlo experiments based on 5000 repetitions. Standard deviations in brackets.

Table 3
Empirical coverage probability for a nominal 95% confidence interval. Location shift model.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100				
0.25	25	0.960	0.972	0.956	0.951	0.994	0.996	0.983	0.977	0.974	0.977	0.944	0.930
	50	0.951	0.967	0.957	0.951	0.992	0.995	0.975	0.963	0.967	0.975	0.940	0.936
	100	0.945	0.960	0.953	0.953	0.990	0.992	0.970	0.951	0.959	0.967	0.940	0.943
	200	0.949	0.959	0.953	0.947	0.988	0.985	0.956	0.951	0.952	0.966	0.936	0.932
0.50	25	0.979	0.977	0.961	0.956	0.988	0.981	0.949	0.952	0.995	0.995	0.986	0.981
	50	0.973	0.968	0.956	0.957	0.981	0.966	0.948	0.954	0.991	0.989	0.977	0.975
	100	0.967	0.965	0.960	0.959	0.976	0.956	0.948	0.945	0.988	0.986	0.974	0.965
	200	0.966	0.955	0.946	0.939	0.971	0.941	0.930	0.944	0.986	0.979	0.968	0.969
0.75	25	0.959	0.968	0.952	0.955	0.927	0.938	0.931	0.939	0.977	0.981	0.941	0.932
	50	0.953	0.965	0.947	0.952	0.910	0.938	0.938	0.938	0.965	0.975	0.939	0.942
	100	0.945	0.959	0.952	0.948	0.912	0.938	0.941	0.941	0.960	0.967	0.942	0.935
	200	0.938	0.963	0.959	0.955	0.910	0.947	0.942	0.946	0.957	0.964	0.944	0.941

Notes: Monte Carlo experiments based on 5000 repetitions.

the χ^2_3 case (in the latter for the third quartile) and $T = 5, 10$. Moreover, the bias is much larger for the $\gamma = 1$ case than that for $\gamma = 0.5$.[9] These results suggest that the FE-QR estimator performs well in small samples for the location shift model but may have a large bias for the location-scale shift model where the quantile of interest is evaluated at an associated low density (i.e., $F = \chi^2_3$ and $\tau = 0.75$ case) when T is small. Overall, these simulations confirm that the bias exists for small T and does not depend on n.

4.2. Inference

To study the inference procedure based on the FE-QR estimator, we first compute the estimated standard error.[10] The results are reported in Tables 2, 5 and 8 for the location shift and location-scale shift ($\gamma = 0.5, 1$) cases, respectively. We also report the sample standard deviation of the estimator based on the Monte Carlo repetitions.

By comparing Table 2 with 1, 5 with 4 and 8 with 7, we may see that the estimated standard error approximates very closely the truth. Second, we calculate the empirical coverage probability of the asymptotic Gaussian confidence interval at the 95% nominal level. In this case, the greater distortions appear in the location-scale shift case for large n/T, and in particular for the χ^2_3 case and $\tau = 0.75$. The distortion is very severe for $T = 5, 10$ and $n = 200$ for all distributions, despite the fact that the estimated standard error approximates well the truth. This possibly reflects that the variance of the FE-QR estimator decreases when nT increases while the bias decreases when T increases but is independent of n, so that the centering of the confidence interval will be severely distorted when n/T is large.

5. Extension: dynamic case

We now extend the asymptotic results in Section 3 to the dynamic case where we allow for dependence across time

9 Although not reported, we have also performed the same experiments for $\gamma = 0.2$. In this case the bias is smaller than that for $\gamma = 0.5$.

10 For estimation of the asymptotic covariance matrix, we use the Gaussian kernel and the default bandwidth option in the `quantreg` package in R.

Table 4

Bias of $\hat{\beta}(\tau)$. Location-scale shift model. $\gamma = 0.5$.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0, 1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100	5	10	50	100
0.25	25	0.068	0.031	0.006	0.002	0.057	0.018	0.003	0.000	0.173	0.105	0.024	0.012
		[0.205]	[0.137]	[0.06]	[0.043]	[0.287]	[0.188]	[0.079]	[0.056]	[0.426]	[0.269]	[0.119]	[0.083]
	50	0.063	0.032	0.006	0.003	0.040	0.014	0.000	0.000	0.181	0.114	0.026	0.013
		[0.148]	[0.098]	[0.042]	[0.03]	[0.203]	[0.131]	[0.056]	[0.039]	[0.283]	[0.186]	[0.084]	[0.059]
	100	0.062	0.029	0.006	0.002	0.034	0.007	0.001	0.001	0.193	0.117	0.028	0.014
		[0.105]	[0.07]	[0.03]	[0.021]	[0.145]	[0.092]	[0.039]	[0.028]	[0.195]	[0.131]	[0.059]	[0.041]
	200	0.063	0.030	0.005	0.002	0.026	0.006	0.000	−0.001	0.192	0.120	0.025	0.014
		[0.073]	[0.048]	[0.021]	[0.014]	[0.101]	[0.066]	[0.028]	[0.02]	[0.136]	[0.093]	[0.041]	[0.03]
0.50	25	0.000	−0.001	−0.002	−0.001	−0.045	−0.028	−0.005	−0.005	0.002	0.001	−0.002	0.000
		[0.185]	[0.13]	[0.055]	[0.039]	[0.377]	[0.27]	[0.117]	[0.082]	[0.293]	[0.182]	[0.07]	[0.049]
	50	0.001	0.000	0.001	0.000	−0.059	−0.030	−0.007	−0.004	−0.006	0.002	−0.001	0.000
		[0.133]	[0.092]	[0.039]	[0.027]	[0.266]	[0.19]	[0.083]	[0.056]	[0.198]	[0.126]	[0.051]	[0.035]
	100	0.001	−0.002	0.000	0.000	−0.069	−0.038	−0.006	−0.004	0.003	−0.001	0.000	−0.001
		[0.095]	[0.065]	[0.027]	[0.019]	[0.189]	[0.132]	[0.058]	[0.041]	[0.138]	[0.088]	[0.035]	[0.025]
	200	0.000	0.000	0.001	0.000	−0.072	−0.036	−0.008	−0.004	0.000	0.000	−0.001	0.000
		[0.065]	[0.046]	[0.02]	[0.014]	[0.134]	[0.095]	[0.042]	[0.029]	[0.097]	[0.063]	[0.025]	[0.017]
0.75	25	−0.066	−0.034	−0.006	−0.004	−0.253	−0.132	−0.023	−0.014	−0.175	−0.105	−0.027	−0.013
		[0.209]	[0.136]	[0.061]	[0.042]	[0.578]	[0.396]	[0.181]	[0.127]	[0.424]	[0.271]	[0.119]	[0.083]
	50	−0.067	−0.031	−0.005	−0.003	−0.263	−0.140	−0.027	−0.014	−0.192	−0.112	−0.028	−0.013
		[0.146]	[0.097]	[0.042]	[0.03]	[0.401]	[0.281]	[0.128]	[0.088]	[0.282]	[0.188]	[0.083]	[0.059]
	100	−0.063	−0.031	−0.005	−0.003	−0.276	−0.148	−0.026	−0.014	−0.192	−0.119	−0.026	−0.015
		[0.104]	[0.069]	[0.03]	[0.021]	[0.286]	[0.2]	[0.09]	[0.064]	[0.195]	[0.131]	[0.059]	[0.042]
	200	−0.062	−0.032	−0.005	−0.002	−0.279	−0.146	−0.026	−0.014	−0.198	−0.117	−0.028	−0.014
		[0.073]	[0.048]	[0.021]	[0.014]	[0.204]	[0.141]	[0.065]	[0.047]	[0.135]	[0.091]	[0.042]	[0.029]

Notes: Monte Carlo experiments based on 5000 repetitions. Standard deviations in brackets.

Table 5

Estimated standard errors of $\hat{\beta}(\tau)$. Location-scale shift model. $\gamma = 0.5$.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0, 1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100	5	10	50	100
0.25	25	0.212	0.147	0.061	0.043	0.378	0.244	0.090	0.061	0.423	0.285	0.113	0.080
		[0.055]	[0.028]	[0.007]	[0.004]	[0.091]	[0.042]	[0.009]	[0.005]	[0.147]	[0.068]	[0.016]	[0.009]
	50	0.145	0.101	0.043	0.030	0.246	0.160	0.061	0.042	0.277	0.194	0.080	0.056
		[0.03]	[0.015]	[0.004]	[0.002]	[0.046]	[0.022]	[0.005]	[0.002]	[0.072]	[0.036]	[0.009]	[0.005]
	100	0.101	0.071	0.030	0.021	0.164	0.107	0.042	0.029	0.187	0.134	0.057	0.040
		[0.016]	[0.009]	[0.002]	[0.001]	[0.024]	[0.011]	[0.002]	[0.001]	[0.035]	[0.019]	[0.005]	[0.003]
	200	0.071	0.049	0.021	0.015	0.111	0.073	0.029	0.020	0.129	0.094	0.040	0.028
		[0.009]	[0.005]	[0.001]	[0.001]	[0.013]	[0.006]	[0.001]	[0.001]	[0.019]	[0.01]	[0.003]	[0.002]
0.50	25	0.212	0.144	0.058	0.040	0.430	0.290	0.119	0.083	0.385	0.235	0.083	0.055
		[0.052]	[0.025]	[0.006]	[0.003]	[0.112]	[0.055]	[0.012]	[0.007]	[0.107]	[0.044]	[0.008]	[0.004]
	50	0.144	0.098	0.040	0.028	0.291	0.196	0.083	0.058	0.249	0.153	0.056	0.038
		[0.028]	[0.014]	[0.003]	[0.002]	[0.06]	[0.029]	[0.007]	[0.004]	[0.052]	[0.022]	[0.004]	[0.002]
	100	0.099	0.068	0.028	0.020	0.201	0.135	0.058	0.041	0.166	0.103	0.038	0.026
		[0.015]	[0.008]	[0.002]	[0.001]	[0.032]	[0.016]	[0.004]	[0.002]	[0.026]	[0.011]	[0.002]	[0.001]
	200	0.069	0.047	0.020	0.014	0.140	0.094	0.040	0.029	0.112	0.070	0.026	0.018
		[0.008]	[0.004]	[0.001]	[0.001]	[0.017]	[0.009]	[0.002]	[0.001]	[0.013]	[0.006]	[0.001]	[0.001]
0.75	25	0.211	0.147	0.061	0.043	0.512	0.384	0.174	0.124	0.421	0.285	0.114	0.080
		[0.055]	[0.029]	[0.007]	[0.004]	[0.164]	[0.087]	[0.025]	[0.014]	[0.143]	[0.07]	[0.017]	[0.009]
	50	0.144	0.102	0.043	0.030	0.361	0.271	0.123	0.088	0.277	0.194	0.080	0.056
		[0.03]	[0.015]	[0.004]	[0.002]	[0.089]	[0.048]	[0.013]	[0.008]	[0.068]	[0.037]	[0.009]	[0.005]
	100	0.100	0.071	0.030	0.021	0.254	0.193	0.087	0.062	0.186	0.134	0.057	0.040
		[0.016]	[0.008]	[0.002]	[0.001]	[0.049]	[0.026]	[0.008]	[0.004]	[0.034]	[0.019]	[0.005]	[0.003]
	200	0.071	0.050	0.021	0.015	0.180	0.137	0.062	0.044	0.128	0.094	0.040	0.028
		[0.009]	[0.005]	[0.001]	[0.001]	[0.027]	[0.015]	[0.004]	[0.003]	[0.018]	[0.01]	[0.003]	[0.002]

Notes: Monte Carlo experiments based on 5000 repetitions. Standard deviations in brackets.

while maintaining independence across individuals. We make the following assumptions in this case.

(D1) $\{(y_{it}, \boldsymbol{x}_{it}), t \geq 1\}$ is stationary and β-mixing for each fixed i, and independent across i. Let $\beta_i(j)$ denote the β-mixing coefficients of $\{(y_{it}, \boldsymbol{x}_{it}), t \geq 1\}$. Then, there exist constants $a \in (0, 1)$ and $B > 0$ such that $\sup_{i \geq 1} \beta_i(j) \leq Ba^j$ for all $j \geq 1$.

(D2) Let $f_{i,j}(u_1, u_{1+j} | \boldsymbol{x}_1, \boldsymbol{x}_{1+j})$ denote the conditional density of $(u_{i1}, u_{i,1+j})$ given $(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i,1+j}) = (\boldsymbol{x}_1, \boldsymbol{x}_{1+j})$. There exists a constant $C_f' > 0$ such that $f_{i,j}(u_1, u_{1+j} | \boldsymbol{x}_1, \boldsymbol{x}_{1+j}) \leq C_f'$ uniformly over $(u_1, u_{1+j}, \boldsymbol{x}_1, \boldsymbol{x}_{1+j})$ for all $i \geq 1$ and $j \geq 1$.

(D3) Let $\tilde{V}_{ni}$ denote the covariance matrix of the term $T^{-1/2} \sum_{t=1}^{T} \{\tau - I(u_{it} \leq 0)\}(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_{it})$. Then, the limit $\tilde{V} := n^{-1} \sum_{i=1}^{n} \tilde{V}_{ni}$ exists and is nonsingular.

Condition (D1) is similar to Condition 1 of Hahn and Kuersteiner (2004). Condition (D2) imposes new restrictions on the conditional densities. Note that in Condition (D3) $\tilde{V}_{ni}$ is now a long run covariance matrix.

The next theorem shows that similar asymptotic results to those in Section 3 are obtained for the dependent case. The proof is given in the Appendix.

Table 6
Empirical coverage probability for a nominal 95% confidence interval. Location-scale shift model. $\gamma = 0.5$.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100	5	10	50	100
0.25	25	0.913	0.946	0.950	0.947	0.976	0.984	0.970	0.970	0.894	0.917	0.913	0.929
	50	0.899	0.933	0.946	0.946	0.974	0.976	0.969	0.959	0.847	0.884	0.908	0.924
	100	0.875	0.931	0.942	0.949	0.967	0.974	0.963	0.951	0.764	0.830	0.903	0.924
	200	0.836	0.904	0.939	0.955	0.960	0.966	0.954	0.953	0.633	0.726	0.886	0.906
0.50	25	0.956	0.957	0.957	0.949	0.945	0.940	0.950	0.952	0.983	0.987	0.979	0.974
	50	0.952	0.960	0.948	0.954	0.939	0.941	0.946	0.955	0.980	0.979	0.968	0.966
	100	0.948	0.957	0.958	0.958	0.933	0.929	0.943	0.944	0.978	0.976	0.967	0.959
	200	0.960	0.955	0.939	0.939	0.915	0.924	0.937	0.938	0.976	0.967	0.949	0.956
0.75	25	0.906	0.940	0.942	0.950	0.826	0.896	0.922	0.936	0.893	0.915	0.917	0.928
	50	0.900	0.932	0.946	0.942	0.807	0.880	0.931	0.943	0.841	0.889	0.912	0.923
	100	0.876	0.926	0.941	0.942	0.735	0.845	0.925	0.937	0.758	0.825	0.906	0.916
	200	0.840	0.900	0.944	0.955	0.623	0.789	0.913	0.929	0.616	0.739	0.870	0.907

Notes: Monte Carlo experiments based on 5000 repetitions.

Table 7
Bias of $\hat{\beta}(\tau)$. Location-scale shift model. $\gamma = 1$.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100	5	10	50	100
0.25	25	0.138	0.066	0.012	0.005	0.129	0.043	0.005	0.000	0.340	0.212	0.049	0.025
		[0.348]	[0.231]	[0.1]	[0.071]	[0.498]	[0.32]	[0.132]	[0.092]	[0.715]	[0.45]	[0.199]	[0.138]
	50	0.132	0.067	0.012	0.006	0.100	0.037	0.000	0.000	0.361	0.224	0.053	0.027
		[0.253]	[0.164]	[0.069]	[0.05]	[0.354]	[0.224]	[0.092]	[0.065]	[0.473]	[0.313]	[0.14]	[0.097]
	100	0.131	0.062	0.013	0.005	0.087	0.025	0.001	0.001	0.380	0.231	0.056	0.028
		[0.178]	[0.118]	[0.049]	[0.035]	[0.253]	[0.156]	[0.065]	[0.047]	[0.326]	[0.219]	[0.098]	[0.068]
	200	0.132	0.065	0.010	0.005	0.074	0.023	0.000	−0.001	0.379	0.237	0.052	0.029
		[0.12]	[0.08]	[0.035]	[0.024]	[0.176]	[0.112]	[0.047]	[0.032]	[0.228]	[0.155]	[0.069]	[0.05]
0.50	25	0.000	−0.002	−0.003	−0.001	−0.079	−0.052	−0.012	−0.009	0.001	0.001	−0.003	0.000
		[0.313]	[0.218]	[0.091]	[0.064]	[0.649]	[0.454]	[0.194]	[0.135]	[0.488]	[0.302]	[0.116]	[0.081]
	50	0.002	0.000	0.001	0.000	−0.103	−0.057	−0.014	−0.008	−0.009	0.002	−0.002	0.000
		[0.227]	[0.153]	[0.065]	[0.045]	[0.456]	[0.32]	[0.137]	[0.093]	[0.331]	[0.21]	[0.084]	[0.058]
	100	0.002	−0.002	0.000	−0.001	−0.123	−0.071	−0.014	−0.008	0.004	−0.002	0.000	−0.001
		[0.162]	[0.108]	[0.044]	[0.032]	[0.326]	[0.223]	[0.096]	[0.068]	[0.232]	[0.147]	[0.058]	[0.041]
	200	−0.001	0.000	0.001	0.000	−0.125	−0.067	−0.016	−0.009	−0.001	0.001	−0.002	0.000
		[0.11]	[0.074]	[0.033]	[0.023]	[0.231]	[0.16]	[0.069]	[0.048]	[0.162]	[0.104]	[0.041]	[0.029]
0.75	25	−0.135	−0.071	−0.012	−0.007	−0.506	−0.271	−0.051	−0.028	−0.344	−0.210	−0.056	−0.027
		[0.356]	[0.229]	[0.101]	[0.07]	[0.986]	[0.668]	[0.299]	[0.21]	[0.71]	[0.453]	[0.196]	[0.138]
	50	−0.139	−0.066	−0.011	−0.006	−0.522	−0.285	−0.055	−0.028	−0.379	−0.224	−0.056	−0.027
		[0.248]	[0.164]	[0.071]	[0.05]	[0.683]	[0.472]	[0.212]	[0.146]	[0.472]	[0.314]	[0.138]	[0.098]
	100	−0.132	−0.066	−0.013	−0.006	−0.546	−0.296	−0.054	−0.028	−0.380	−0.235	−0.054	−0.030
		[0.178]	[0.116]	[0.05]	[0.036]	[0.487]	[0.338]	[0.15]	[0.106]	[0.33]	[0.221]	[0.098]	[0.07]
	200	−0.126	−0.068	−0.011	−0.004	−0.550	−0.293	−0.055	−0.028	−0.390	−0.231	−0.056	−0.028
		[0.123]	[0.08]	[0.035]	[0.024]	[0.348]	[0.237]	[0.107]	[0.077]	[0.227]	[0.153]	[0.069]	[0.047]

Notes: Monte Carlo experiments based on 5000 repetitions. Standard deviations in brackets.

Theorem 5.1. *Assume conditions* (D1)–(D3), (A3) *and* (B1)–(B3). *Then,* $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *is weakly consistent provided that* $(\log n)^2/T \to 0$. *If* $(\log n)^2/T \to 0$ *but T grows at most polynomially in n, then* $\hat{\boldsymbol{\beta}}$ *admits the expansion* (3.2). *If moreover* $n^2(\log n)^3/T \to 0$, *then we have* $\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \Gamma^{-1}\tilde{V}\Gamma^{-1})$.

In proving Theorem 5.1, we need some extensions of empirical process inequalities to β-mixing sequences, which we believe is a nontrivial task. We develop those extensions in Appendix C, which are useful in other contexts such as asymptotic analysis of sieve estimation for β-mixing sequences.

6. Discussion

In this paper, we have studied the asymptotic properties of the FE-QR estimator. The results found in this paper show that the asymptotic theory for panel models with non-differentiable objective functions, as in the QR case, should be analyzed carefully. Usually the limiting distribution under the joint asymptotics coincides with that under the sequential asymptotics as long as n/T goes to zero, as is well recognized in the literature. However, this paper draws a caution that such a result may not directly apply to the QR case.

There remain several issues to be investigated.

It is an open question whether the convergence rate of the remainder term in (3.2) can be improved to $O_p(T^{-1})$. It should be pointed out that although the rate of the remainder term derived in this paper is the best one that we could achieve at this point, there might be a room for improvement on the rate, which means that our condition for the asymptotic normality is only a sufficient one. However, although we could not formally show in this paper, we conjecture that $n/T \to 0$ is a sufficient condition to asymptotic normality of QR panel data. Kato and Galvao (2010) used a smoothed version of the FE-QR estimator to derive the asymptotic bias of the estimator when $n/T \to \rho > 0$. Thus, the smoothed estimator is unbiased for $n/T \to 0$. However, it is important to note that the derivation makes use of the smoothness of the objective and the score functions, which is not applicable to this paper. The challenge in the present context is that higher order expansions for the standard QR is a very difficult subject.

Table 8
Estimated standard errors of $\hat{\beta}(\tau)$. Location-scale shift model. $\gamma = 1$.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0, 1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi_3^2$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100	5	10	50	100
0.25	25	0.362	0.249	0.102	0.071	0.626	0.403	0.151	0.102	0.710	0.480	0.191	0.134
		[0.099]	[0.05]	[0.012]	[0.007]	[0.155]	[0.073]	[0.015]	[0.008]	[0.255]	[0.116]	[0.028]	[0.015]
	50	0.248	0.171	0.071	0.050	0.411	0.267	0.103	0.070	0.466	0.327	0.134	0.094
		[0.053]	[0.027]	[0.007]	[0.004]	[0.081]	[0.038]	[0.008]	[0.004]	[0.124]	[0.063]	[0.015]	[0.009]
	100	0.173	0.119	0.050	0.035	0.276	0.180	0.070	0.048	0.315	0.225	0.095	0.067
		[0.029]	[0.015]	[0.004]	[0.002]	[0.042]	[0.02]	[0.004]	[0.002]	[0.061]	[0.033]	[0.008]	[0.005]
	200	0.120	0.083	0.035	0.025	0.189	0.123	0.049	0.034	0.217	0.158	0.067	0.047
		[0.015]	[0.008]	[0.002]	[0.001]	[0.022]	[0.011]	[0.002]	[0.001]	[0.032]	[0.018]	[0.005]	[0.003]
0.50	25	0.359	0.243	0.097	0.067	0.730	0.493	0.201	0.139	0.632	0.387	0.138	0.092
		[0.094]	[0.045]	[0.01]	[0.006]	[0.2]	[0.099]	[0.022]	[0.012]	[0.18]	[0.076]	[0.014]	[0.007]
	50	0.245	0.166	0.067	0.047	0.498	0.337	0.139	0.097	0.411	0.254	0.093	0.063
		[0.05]	[0.025]	[0.006]	[0.003]	[0.108]	[0.053]	[0.012]	[0.007]	[0.089]	[0.038]	[0.008]	[0.004]
	100	0.170	0.115	0.047	0.033	0.345	0.232	0.097	0.068	0.275	0.171	0.064	0.043
		[0.027]	[0.014]	[0.003]	[0.002]	[0.057]	[0.029]	[0.006]	[0.004]	[0.045]	[0.02]	[0.004]	[0.002]
	200	0.117	0.080	0.033	0.023	0.241	0.162	0.067	0.048	0.186	0.116	0.044	0.030
		[0.014]	[0.007]	[0.002]	[0.001]	[0.031]	[0.016]	[0.004]	[0.002]	[0.023]	[0.01]	[0.002]	[0.001]
0.75	25	0.360	0.249	0.103	0.071	0.892	0.660	0.290	0.205	0.706	0.479	0.191	0.134
		[0.098]	[0.051]	[0.012]	[0.007]	[0.297]	[0.155]	[0.042]	[0.023]	[0.246]	[0.121]	[0.028]	[0.016]
	50	0.247	0.172	0.071	0.050	0.629	0.463	0.204	0.145	0.466	0.327	0.135	0.094
		[0.053]	[0.027]	[0.007]	[0.004]	[0.162]	[0.084]	[0.023]	[0.013]	[0.119]	[0.064]	[0.015]	[0.009]
	100	0.172	0.120	0.050	0.035	0.442	0.328	0.145	0.103	0.314	0.225	0.095	0.067
		[0.028]	[0.015]	[0.004]	[0.002]	[0.087]	[0.047]	[0.013]	[0.007]	[0.06]	[0.033]	[0.008]	[0.005]
	200	0.121	0.084	0.035	0.025	0.313	0.233	0.103	0.073	0.215	0.158	0.067	0.047
		[0.016]	[0.008]	[0.002]	[0.001]	[0.047]	[0.026]	[0.007]	[0.004]	[0.032]	[0.018]	[0.004]	[0.003]

Notes: Monte Carlo experiments based on 5000 repetitions. Standard deviations in brackets.

Table 9
Empirical coverage probability for a nominal 95% confidence interval. Location-scale shift model. $\gamma = 1$.

τ	n/T	$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0, 1)$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi_3^2$				$\epsilon_{it} \overset{\text{i.i.d.}}{\sim}$ *Cauchy*			
		5	10	50	100	5	10	50	100	5	10	50	100
0.25	25	0.907	0.939	0.948	0.949	0.971	0.978	0.971	0.973	0.875	0.900	0.913	0.929
	50	0.883	0.929	0.947	0.944	0.966	0.974	0.970	0.963	0.812	0.863	0.905	0.923
	100	0.857	0.920	0.945	0.950	0.957	0.971	0.966	0.958	0.714	0.792	0.888	0.916
	200	0.783	0.883	0.939	0.953	0.949	0.963	0.948	0.964	0.546	0.651	0.857	0.882
0.50	25	0.956	0.961	0.956	0.953	0.940	0.937	0.950	0.956	0.979	0.985	0.979	0.974
	50	0.949	0.959	0.949	0.956	0.937	0.939	0.949	0.957	0.979	0.978	0.967	0.967
	100	0.949	0.958	0.963	0.950	0.926	0.930	0.944	0.946	0.977	0.975	0.968	0.960
	200	0.959	0.965	0.939	0.941	0.912	0.922	0.941	0.940	0.974	0.970	0.959	0.960
0.75	25	0.896	0.938	0.945	0.950	0.812	0.892	0.924	0.935	0.873	0.899	0.918	0.927
	50	0.884	0.926	0.945	0.945	0.790	0.863	0.926	0.940	0.813	0.867	0.903	0.919
	100	0.847	0.917	0.936	0.935	0.697	0.819	0.922	0.932	0.703	0.781	0.892	0.910
	200	0.812	0.864	0.931	0.956	0.552	0.737	0.903	0.921	0.535	0.664	0.846	0.903

Notes: Monte Carlo experiments based on 5000 repetitions.

Since there is a large literature on analytical bias correction for large panel data, one could wonder about deriving the asymptotic bias in the present context of FE-QR estimation. There are at least two important reasons to explain the degree of difficulty in the FE-QR case. First, the rate $O_p\{(T/\log n)^{-3/4}\}$ in the Bahadur representation in Theorem 3.2 comes from the rate of the score terms, as defined in the proof of Theorem 3.2. Unfortunately, a direct expansion of these terms with respect to $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and the simple evaluation of the mean and variance is not feasible.[11] It is important to note that for each i, the convergence rate of $(\hat{\alpha}_i, \hat{\boldsymbol{\beta}})$ is dominated by $\hat{\alpha}_i$, and thus is at most $T^{-1/2}$. However, because of the non-smoothness of the indicator function, the evaluation of these terms based on some moment inequalities for empirical processes (such as Proposition B.1) leads to the rate $O_p\{\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}|^{3/2}\}$, which turns out to be $O_p(T^{-3/4})$ ($\log n$ term is ignored for simplicity). Thus, a more refined result (such as a bias result) could be obtained if one could establish the probability limits of these terms (scaled by a suitable term), which is thought to be a quite challenging task and is not solved in this paper.[12] Second, there is another difficulty to obtain a bias result to the FE-QR estimator. This is related to indeterminateness of the higher order behavior of quantile regression estimators. Consider, for illustrative purposes, a sample τ-quantile of uniform random variables $u_1, \ldots, u_n$ on [0, 1] where $\tau \in (0, 1)$ is fixed and $n\tau$ is an integer. Let $u_{(1)} < \cdots < u_{(n)}$ denote the order statistics of $u_1, \ldots, u_n$. Then, the sample τ-quantile is usually given by $u_{(n\tau)}$. However, if we view the sample quantile as a solution to the QR minimization problem, it can be any value in $[u_{(n\tau)}, u_{(n\tau+1)}]$, of which the mean length is of order n^{-1}. This means that the higher order behavior of the sample τ-quantile at n^{-1} rate is not fully determined if we take the sample

11 A way to deal with such terms is to consider them as empirical processes indexed by $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and establish the rates by using the preliminary rates of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. This is what the present proof does.

12 To obtain a bias result, establishing the exact probability limits of these terms would be essential, because the corresponding terms in the standard smooth case contribute to the bias of the resulting fixed effects estimator.

quantile as a solution to the QR minimization problem. Since the asymptotic bias of the general fixed effect estimator depends on the higher order behavior of the estimators of the individual parameters at T^{-1} rate, this indeterminateness would be another challenge to obtain a bias result to the FE-QR estimator.

Appendix A. Proofs

A.1. Proof of Theorem 3.1

Put $\mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) := T^{-1}\sum_{t=1}^{T}\rho_\tau(y_{it}-\alpha_i-\boldsymbol{x}_{it}'\boldsymbol{\beta})$ and $\Delta_{ni}(\alpha_i, \boldsymbol{\beta}) := \mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)$. For each $\delta > 0$, define $B_i(\delta) := \{(\alpha, \boldsymbol{\beta}): |\alpha - \alpha_{i0}| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 \le \delta\}$ and $\partial B_i(\delta) := \{(\alpha, \boldsymbol{\beta}): |\alpha - \alpha_{i0}| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 = \delta\}$.

Proof of Theorem 3.1. We divide the proof into two steps.

Step 1. We first prove $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. Fix any $\delta > 0$. For each $(\alpha_i, \boldsymbol{\beta}) \notin B_i(\delta)$, define $\tilde{\alpha}_i = r_i\alpha_i + (1 - r_i)\alpha_{i0}$, $\tilde{\boldsymbol{\beta}}_i = r_i\boldsymbol{\beta} + (1 - r_i)\boldsymbol{\beta}_0$, where $r_i = \delta/(|\alpha_i - \alpha_{i0}| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1)$. Note that $r_i \in (0, 1)$ and $(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}_i) \in \partial B_i(\delta)$. Because of the convexity of the objective function, we have

$$\begin{aligned} &r_i\{\mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)\} \\ &\quad\ge \mathbb{M}_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0) \\ &\quad= \{E[\Delta_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}_i)]\} + \{\Delta_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}_i) - E[\Delta_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}_i)]\}. \end{aligned} \tag{A.1}$$

Use the identity of Knight (1998) to obtain

$$E[\Delta_{ni}(\alpha_i, \boldsymbol{\beta})] = E\left[\int_0^{(\alpha_i-\alpha_{i0})+\boldsymbol{x}_{i1}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}\{F_i(s|\boldsymbol{x}_{i1}) - \tau\}ds\right].$$

From condition (A3), the first term on the right side of Eq. (A.1) is greater than or equal to ϵ_δ for all $1 \le i \le n$. Thus, by (A.1), we obtain the inclusion relation

$$\begin{aligned} &\left\{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 > \delta\right\} \\ &\quad\subset \left\{\mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) \le \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0),\ 1 \le \exists i \le n,\ \exists(\alpha_i, \boldsymbol{\beta}) \notin B_i(\delta)\right\} \\ &\quad\subset \left\{\max_{1\le i\le n}\sup_{(\alpha_i,\boldsymbol{\beta})\in B_i(\delta)}|\Delta_{ni}(\alpha_i, \boldsymbol{\beta}) - E[\Delta_{ni}(\alpha_i, \boldsymbol{\beta})]| \ge \epsilon_\delta\right\}. \end{aligned}$$

The first inclusion follows from the following argument. Suppose that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 > \delta$. Then, $(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) \notin B_i(\delta)$ for all $1 \le i \le n$. If $\mathbb{M}_{ni}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) > \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)$ for all $1 \le i \le n$, then $\sum_{i=1}^{n}\mathbb{M}_{ni}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) > \sum_{i=1}^{n}\mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)$, which however contradicts the definition of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. Thus, $\mathbb{M}_{ni}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) \le \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)$ for some $1 \le i \le n$, which leads to the first inclusion.

Therefore, it suffices to show that for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left\{\max_{1\le i\le n}\sup_{(\alpha_i,\boldsymbol{\beta})\in B_i(\delta)}|\Delta_{ni}(\alpha_i, \boldsymbol{\beta}) - E[\Delta_{ni}(\alpha_i, \boldsymbol{\beta})]| > \epsilon\right\} = 0. \tag{A.2}$$

[Recall that $T = T_n$ is indexed by n, and $n \to \infty$ automatically means that $T = T_n \to \infty$.]

Because of the union bound, it suffices to prove that for every $\epsilon > 0$,

$$\max_{1\le i\le n} P\left\{\sup_{(\alpha,\boldsymbol{\beta})\in B_i(\delta)}|\Delta_{ni}(\alpha, \boldsymbol{\beta}) - E[\Delta_{ni}(\alpha, \boldsymbol{\beta})]| > \epsilon\right\} = o(n^{-1}). \tag{A.3}$$

We follow the proof of Fernandez-Val (2005, Lemma 7) to show (A.3). Without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. Then, $B_i(\delta)$ is independent of i and write $B_i(\delta) = B(\delta)$ for simplicity. Put $g_{\alpha,\boldsymbol{\beta}}(u, \boldsymbol{x}) := \rho_\tau(u - \alpha - \boldsymbol{x}'\boldsymbol{\beta}) - \rho_\tau(u)$. Observe that $|g_{\alpha,\boldsymbol{\beta}}(u, \boldsymbol{x}) - g_{\bar{\alpha},\bar{\boldsymbol{\beta}}}(u, \boldsymbol{x})| \le C(1 + \|\boldsymbol{x}\|_1)(|\alpha - \bar{\alpha}| + \|\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}\|_1)$ for some universal constant $C > 0$. Put $L(\boldsymbol{x}) := C(1 + \|\boldsymbol{x}\|_1)$ and $\kappa := \sup_{i\ge1} E[L(\boldsymbol{x}_{i1})]$. Since $B(\delta)$ is a compact subset of $\mathbb{R}^{p+1}$, there exist K ℓ_1-balls with centers $(\alpha^{(j)}, \boldsymbol{\beta}^{(j)})$, $j = 1, \ldots, K$ and radius $\epsilon/(7\kappa)$ such that the collection of these balls covers $B(\delta)$. Note that K is independent of i and can be chosen such that $K = K(\epsilon) = O(\epsilon^{-p-1})$ as $\epsilon \to 0$. Now, for each $(\alpha, \boldsymbol{\beta}) \in B(\delta)$, there is $j \in \{1, \ldots, K\}$ such that $|g_{\alpha,\boldsymbol{\beta}}(u, \boldsymbol{x}) - g_{\alpha^{(j)},\boldsymbol{\beta}^{(j)}}(u, \boldsymbol{x})| \le L(\boldsymbol{x})\epsilon/(7\kappa)$, which leads to $|\Delta_{ni}(\alpha, \boldsymbol{\beta}) - E[\Delta_{ni}(\alpha, \boldsymbol{\beta})]| \le |\Delta_{ni}(\alpha^{(j)}, \boldsymbol{\beta}^{(j)}) - E[\Delta_{ni}(\alpha^{(j)}, \boldsymbol{\beta}^{(j)})]| + \{\epsilon/(7\kappa)\}\cdot|T^{-1}\sum_{t=1}^{T}\{L(\boldsymbol{x}_{it}) - E[L(\boldsymbol{x}_{i1})]\}| + 2\epsilon/7$. Therefore, we have

$$\begin{aligned} &P\left\{\sup_{(\alpha,\boldsymbol{\beta})\in B(\delta)}|\Delta_{ni}(\alpha, \boldsymbol{\beta}) - E[\Delta_{ni}(\alpha, \boldsymbol{\beta})]| > \epsilon\right\} \\ &\quad\le \sum_{j=1}^{K} P\left\{|\Delta_{ni}(\alpha^{(j)}, \boldsymbol{\beta}^{(j)}) - E[\Delta_{ni}(\alpha^{(j)}, \boldsymbol{\beta}^{(j)})]| > \frac{\epsilon}{3}\right\} \\ &\qquad + P\left\{\frac{1}{T}\left|\sum_{t=1}^{T}\{L(\boldsymbol{x}_{it}) - E[L(\boldsymbol{x}_{i1})]\}\right| > \frac{7\kappa}{3}\right\}. \end{aligned} \tag{A.4}$$

Since $\sup_{i\ge1} E[L^{2s}(\boldsymbol{x}_{i1})] < \infty$, application of the Marcinkiewicz–Zygmund inequality (see Corollary 2 in Chow and Teicher (1997, p. 387)) implies that both terms on the right side of (A.4) are $O(T^{-s})$ uniformly over $1 \le i \le n$. Because of the hypothesis on T, they are $o(n^{-1})$, leading to (A.3).

Step 2. Next, we shall show that $\max_{1\le i\le n}|\hat{\alpha}_i - \alpha_{i0}| \xrightarrow{p} 0$. Recall that $\hat{\alpha}_i = \arg\min_\alpha \mathbb{M}_{ni}(\alpha, \hat{\boldsymbol{\beta}})$. Fix any $\delta > 0$. For each $\alpha_i \in \mathbb{R}$ such that $|\alpha_i - \alpha_{i0}| > \delta$, define $\tilde{\alpha}_i = r_i\alpha_i + (1 - r_i)\alpha_{i0}$, where $r_i = \delta/|\alpha_i - \alpha_{i0}|$. Because of the convexity of the objective function, we have

$$\begin{aligned} &r_i\{\mathbb{M}_{ni}(\alpha_i, \hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0}, \hat{\boldsymbol{\beta}})\} \\ &\quad\ge \mathbb{M}_{ni}(\tilde{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0}, \hat{\boldsymbol{\beta}}) \\ &\quad= \mathbb{M}_{ni}(\tilde{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0) - \{\mathbb{M}_{ni}(\alpha_{i0}, \hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)\} \\ &\quad= \{\Delta_{ni}(\tilde{\alpha}_i, \hat{\boldsymbol{\beta}}) - E[\Delta_{ni}(\tilde{\alpha}_i, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} \\ &\qquad - \{\Delta_{ni}(\alpha_{0i}, \hat{\boldsymbol{\beta}}) - E[\Delta_{ni}(\alpha_{0i}, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} + E[\Delta_{ni}(\tilde{\alpha}_i, \boldsymbol{\beta}_0)] \\ &\qquad + \{E[\Delta_{ni}(\tilde{\alpha}_i, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - E[\Delta_{ni}(\tilde{\alpha}_i, \boldsymbol{\beta}_0)]\} + E[\Delta_{ni}(\alpha_{0i}, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}. \end{aligned}$$

It is seen from condition (A3) that the third term on the right side is greater than or equal to ϵ_δ. Thus, we obtain the inclusion relation

$$\begin{aligned} &\{|\hat{\alpha}_i - \alpha_{i0}| > \delta,\ 1 \le \exists i \le n\} \\ &\quad\subset \{\mathbb{M}_{ni}(\alpha_i, \hat{\boldsymbol{\beta}}) \le \mathbb{M}_{ni}(\alpha_{i0}, \hat{\boldsymbol{\beta}}),\ 1 \le \exists i \le n, \\ &\quad\ \exists\alpha_i \in \mathbb{R} \text{ s.t. } |\alpha_i - \alpha_{i0}| > \delta\} \\ &\quad\subset \left\{\max_{1\le i\le n}\sup_{|\alpha-\alpha_{i0}|\le\delta}|\Delta_{ni}(\alpha, \hat{\boldsymbol{\beta}}) - E[\Delta_{ni}(\alpha, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}| \ge \frac{\epsilon_\delta}{4}\right\} \\ &\quad\cup \left\{\max_{1\le i\le n}\sup_{|\alpha-\alpha_{i0}|\le\delta}||E[\Delta_{ni}(\alpha, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - E[\Delta_{ni}(\alpha, \boldsymbol{\beta}_0)]| \ge \frac{\epsilon_\delta}{4}\right\} \\ &\quad=: A_{1n} \cup A_{2n}. \end{aligned}$$

Since $\hat{\boldsymbol{\beta}}$ is consistent by Step 1, and especially $\hat{\boldsymbol{\beta}} = O_p(1)$, by (A.2), it is shown that $P(A_{1n}) \to 0$. Finally, since

$$|E[\Delta_{ni}(\alpha, \boldsymbol{\beta})] - E[\Delta_{ni}(\alpha, \boldsymbol{\beta}_0)]| \le 2E[\|\boldsymbol{x}_{i1}\|]\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|,$$

and $\sup_{i\ge1} E[\|\boldsymbol{x}_{i1}\|] \le 1 + \sup_{i\ge1} E[\|\boldsymbol{x}_{i1}\|^{2s}] < \infty$, consistency of $\hat{\boldsymbol{\beta}}$ implies that $P(A_{2n}) \to 0$. Therefore, we complete the proof. □

Remark A.1. If $\sup_{i\geq 1}\|\boldsymbol{x}_{i1}\| \leq M$ (a.s.) for some constant M, we may take $L(\boldsymbol{x}) \equiv C(1+M)$ and the second term on the right side of (A.4) will vanish. In this case, we can apply Hoeffding's inequality to the first term on the right side of (A.4) and the probability in (A.3) is bounded by $D\exp(-DT)$ for some positive constant D that depends on ϵ but not on i. Therefore, the conclusion of Theorem 3.1 holds when $\log n/T \to 0$ as $n \to \infty$ in this case.

A.2. Proof of Theorem 3.2

Define

$$\mathbb{H}_{ni}^{(1)}(\alpha_i,\boldsymbol{\beta}) := \frac{1}{T}\sum_{t=1}^{T}\{\tau - I(y_{it} \leq \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})\},$$

$$\begin{aligned}H_{ni}^{(1)}(\alpha_i,\boldsymbol{\beta}) &:= E[\mathbb{H}_{ni}^{(1)}(\alpha_i,\boldsymbol{\beta})]\\ &= E[\{\tau - F_i(\alpha_i - \alpha_{i0} + \boldsymbol{x}_{i1}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|\boldsymbol{x}_{i1})\}],\end{aligned}$$

$$\mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\tau - I(y_{it} \leq \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})\}\boldsymbol{x}_{it},$$

$$\begin{aligned}&H_{n}^{(2)}(\boldsymbol{\alpha},\boldsymbol{\beta})\\ &:= E[\mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha},\boldsymbol{\beta})]\\ &= \frac{1}{n}\sum_{i=1}^{n}E[\{\tau - F_i(\alpha_i - \alpha_{i0} + \boldsymbol{x}_{i1}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|\boldsymbol{x}_{i1})\}\boldsymbol{x}_{i1}].\end{aligned}$$

Note that $\mathbb{H}_{ni}^{(1)}(\alpha_i,\boldsymbol{\beta})$ depends on n since T does. The $(n+p)$ dimensional vector of functions $[\mathbb{H}_{n1}^{(1)}(\alpha_1,\boldsymbol{\beta}),\ldots,\mathbb{H}_{nn}^{(1)}(\alpha_n,\boldsymbol{\beta}), \mathbb{H}_{n}^{(2)'}(\boldsymbol{\alpha},\boldsymbol{\beta})]'$ are called the scores for problem (2.2).

Before starting the proof, we introduce some notation used in empirical process theory. Let $\mathcal{F}$ be a class of measurable functions on a measurable space $(S,\mathcal{S})$. For a process $Z(f)$ defined on $\mathcal{F}$, $\|Z(f)\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}}|Z(f)|$. For a probability measure Q on $(S,\mathcal{S})$ and $\epsilon > 0$, let $N(\mathcal{F}, L_2(Q), \epsilon)$ denote the ϵ-covering number of $\mathcal{F}$ with respect to the $L_2(Q)$ norm $\|\cdot\|_{L_2(Q)}$. For the definition of a Vapnik–Chervonenkis (VC) subgraph class, we refer to van der Vaart and Wellner (1996, Section 2.6). For $a,b \in \mathbb{R}$, we use the notation $a \vee b := \max\{a,b\}$.

Proof of Theorem 3.2. Recall first that by Theorem 3.1, under the present conditions, $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ is weakly consistent. We divide the proof into several steps.

Step 1 (Asymptotic representation). We shall show that

$$\begin{aligned}&\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + o_p(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|)\\ &= \Gamma_n^{-1}\left\{-n^{-1}\sum_{i=1}^{n}\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\right\}\\ &\quad - \Gamma_n^{-1}\left[n^{-1}\sum_{i=1}^{n}\boldsymbol{\gamma}_i\left\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}})\right.\right.\\ &\quad \left.\left. - \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\right\}\right]\\ &\quad + \Gamma_n^{-1}\left\{\mathbb{H}_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - H_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\right\}\\ &\quad + O_p\left\{T^{-1} \vee \max_{1\leq i\leq n}|\hat{\alpha}_i - \alpha_{i0}|^2\right\}.\end{aligned} \tag{A.5}$$

Because of the computational property of the QR estimator (see Eq. (3.10) of Gutenbrunner and Jureckova (1992)), it is shown that $\max_{1\leq i\leq n}|\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}})| = O_p(T^{-1})$. Thus, uniformly over $1 \leq i \leq n$, we have

$$\begin{aligned}O_p(T^{-1}) &= \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0) + H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}})\\ &\quad + \left\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\right\}.\end{aligned}$$

Expanding $H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}})$ around $(\alpha_{i0},\boldsymbol{\beta}_0)$, we have

$$\begin{aligned}\hat{\alpha}_i - \alpha_{i0} &= \{f_i(0)\}^{-1}\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0) - \boldsymbol{\gamma}_i'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ &\quad + \{f_i(0)\}^{-1}\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}})\\ &\quad - \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\} + \hat{r}_{ni},\end{aligned} \tag{A.6}$$

where $\max_{1\leq i\leq n}|\hat{r}_{ni}| = O_p\{T^{-1} \vee \max_{1\leq i\leq n}|\hat{\alpha}_i - \alpha_{i0}|^2 \vee \|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2\}$.

Similarly, the computational property of the QR estimator implies that $\|\mathbb{H}_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})\| = O_p\{T^{-1}\max_{1\leq i\leq n,1\leq t\leq T}\|\boldsymbol{x}_{it}\|\} = O_p(T^{-1})$, from which we have

$$\begin{aligned}O_p(T^{-1}) &= \mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) + H_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})\\ &\quad + \{\mathbb{H}_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - H_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\}.\end{aligned} \tag{A.7}$$

Use Taylor's theorem to obtain

$$\begin{aligned}H_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) &= -\frac{1}{n}\sum_{i=1}^{n}E[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}](\hat{\alpha}_i - \alpha_{i0})\\ &\quad - \left\{\frac{1}{n}\sum_{i=1}^{n}E[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}']\right\}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ &\quad + o_p(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|) + O_p\left\{\max_{1\leq i\leq n}|\hat{\alpha}_i - \alpha_{i0}|^2\right\}.\end{aligned} \tag{A.8}$$

Plugging (A.6) into (A.8) leads to

$$\begin{aligned}H_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) &= -\frac{1}{n}\sum_{i=1}^{n}\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\boldsymbol{\gamma}_i - \Gamma_n(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\gamma}_i\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}})\\ &\quad - \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\} + o_p(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|)\\ &\quad + O_p\left\{T^{-1} \vee \max_{1\leq i\leq n}|\hat{\alpha}_i - \alpha_{i0}|^2\right\}.\end{aligned} \tag{A.9}$$

Combining (A.7) and (A.9) yields the desired representation. The remaining steps are devoted to determining the order of the remainder terms in (A.5).

Step 2 (Rates of the remainder terms). Take $\delta_n \to 0$ such that $\max_{1\leq i\leq n}|\hat{\alpha}_i - \alpha_{i0}| \vee \|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| = O_p(\delta_n)$. We shall show that

$$\begin{aligned}&\left\|n^{-1}\sum_{i=1}^{n}\boldsymbol{\gamma}_i\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\}\right\|\\ &\quad = O_p(d_n),\end{aligned} \tag{A.10}$$

$$\|\mathbb{H}_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - H_{n}^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\| = O_p(d_n) \tag{A.11}$$

where $d_n := T^{-1}|\log\delta_n| \vee T^{-1/2}\delta_n^{1/2}|\log\delta_n|^{1/2}$.

We only prove (A.10) since the proof of (A.11) is analogous.[13] Without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. Put $g_{\alpha,\boldsymbol{\beta}}(u,\boldsymbol{x}) := I(u \leq \alpha + \boldsymbol{x}'\boldsymbol{\beta}) - I(u \leq 0)$, $\mathcal{G}_\delta :=$

[13] Although the present proof requires $\boldsymbol{x}_{i1}$ to be bounded, it is possible to use Theorem 2.14.1 of van der Vaart and Wellner (1996) to show (A.11), which only requires that $\sup_{i\geq 1}E[\|\boldsymbol{x}_{i1}\|^2] < \infty$. However, recall that condition (B1) is used to ensure (A.7).

$\{g_{\alpha,\boldsymbol{\beta}}: |\alpha| \le \delta, \|\boldsymbol{\beta}\| \le \delta\}$ and $\xi_{it} := (u_{it}, \boldsymbol{x}_{it})$. Since $\boldsymbol{\gamma}_i$ is bounded over i, it suffices to show that

$$\max_{1\le i\le n} E\left[\left\|\sum_{t=1}^{T}\{g(\xi_{it}) - E[g(\xi_{i1})]\}\right\|_{\mathcal{G}_{\delta_n}}\right] = O(d_n T). \tag{A.12}$$

To this end, we apply Proposition B.1 to the class of functions $\tilde{\mathcal{G}}_{i,\delta_n} := \{g - E[g(\xi_{i1})]: g \in \mathcal{G}_{\delta_n}\}$. Observe that $\tilde{\mathcal{G}}_{i,\delta_n}$ is pointwise measurable and each element of $\tilde{\mathcal{G}}_{i,\delta_n}$ is bounded by 2. Because of Lemmas 2.6.15 and 2.6.18 of van der Vaart and Wellner (1996), the class $\mathcal{G}_\infty := \{g_{\alpha,\boldsymbol{\beta}}: \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^p\}$ is a VC subgraph class. Thus, by Theorem 2.6.7 of van der Vaart and Wellner (1996), the fact that $\tilde{\mathcal{G}}_{i,\delta_n} \subset \{g - E[g(\xi_{i1})]: g \in \mathcal{G}_\infty\}$, and a simple estimate of covering numbers, there exist constants $A \ge 3\sqrt{e}$ and $v \ge 1$ independent of i and n such that $N(\tilde{\mathcal{G}}_{i,\delta_n}, L_2(Q), 2\epsilon) \le (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and every probability measure Q on $\mathbb{R}^{p+1}$. Combining the fact that $E[g_{\alpha,\boldsymbol{\beta}}(\xi_{i1})^2] = E[|F_i(\alpha + \boldsymbol{x}_{i1}'\boldsymbol{\beta}|\boldsymbol{x}_{i1}) - F_i(0|\boldsymbol{x}_{i1})|] \le C_f(|\alpha| + M\|\boldsymbol{\beta}\|)$, one can see that $\tilde{\mathcal{G}}_{i,\delta_n}$ satisfies all the conditions of Proposition B.1 with $U = 2$ and $\sigma^2 = C_f(1+M)\delta_n$, and the constants A and v are independent of i and n. This implies that the left side of (A.12) is $O(d_n T)$.

Step 3 (Preliminary convergence rates). We shall show that

$$\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}| = O_p\{(T/\log n)^{-1/2}\},$$

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = o_p\{(T/\log n)^{-1/2}\}.$$

We first show that $\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}| = O_p\{(T/\log n)^{-1/2}\}$. Because of consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and the result given in Step 2, the second and third terms on the right side of (A.5) is $o_p(T^{-1/2})$, which implies that

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\left\{\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}|^2\right\} + o_p(T^{-1/2}). \tag{A.13}$$

Thus, by (A.6), $\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}|$ is bounded by

$$\text{const.} \times \left\{\max_{1\le i\le n} |\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)| + \max_{1\le i\le n} |\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)|\right\} + o_p(T^{-1/2}),$$

with probability approaching one.

First, observe that for any $K > 0$,

$$P\left\{\max_{1\le i\le n} |\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)| > (T/\log n)^{-1/2}K\right\} \le \sum_{i=1}^{n} P\left\{|\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)| > (T/\log n)^{-1/2}K\right\},$$

and the right side is bounded by $2n^{1-K^2/2}$ by Hoeffding's inequality. This implies that $\max_{1\le i\le n} |\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)| = O_p\{(T/\log n)^{-1/2}\}$.

We next show that

$$\max_{1\le i\le n} |\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)| = o_p\{(T/\log n)^{-1/2}\},$$

which leads to the first result. Without loss of generality, as before, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = 0$. Let $\mathcal{G}_\delta$ and ξ_{it} be the same as those given in Step 2. Because of consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and the union bound, it suffices to show that for every $\epsilon > 0$, there exists a sufficiently small $\delta > 0$ such that

$$\max_{1\le i\le n} P\left\{\left\|\sum_{t=1}^{T}\{g(\xi_{it}) - E[g(\xi_{i1})]\}\right\|_{\mathcal{G}_\delta} > (T\log n)^{1/2}\epsilon\right\} = o(n^{-1}).$$

To this end, we make use of Bousquet's version of Talagrand's inequality (see Proposition B.2 in Appendix B). Fix $\epsilon > 0$. Put $Z_i := \|\sum_{t=1}^{T}\{g(\xi_{it}) - E[g(\xi_{i1})]\}\|_{\mathcal{G}_\delta}$. By Proposition B.2, for all $s > 0$, with probability at least $1 - e^{-s^2}$, we have

$$Z_i \le E[Z_i] + s\sqrt{2\{TC_f(1+M)\delta + 4E[Z_i]\}} + \frac{2s^2}{3}, \tag{A.14}$$

where we have used the fact that each element in $\mathcal{G}_\delta$ is bounded by 1 and $E[g^2(\xi_{i1})] \le C_f(1+M)\delta$ for $g \in \mathcal{G}_\delta$. By Step 2, we have

$$\max_{1\le i\le n} E[Z_i] \le \text{const.} \times (\log|\delta| + T^{1/2}\delta^{1/2}|\log\delta|^{1/2}),$$

where the constant is independent of δ and n. Take $s = \sqrt{2\log n}$. Then, it is seen that there exist a positive constant δ and a positive integer n_0 independent of i and n such that the right side on (A.14) is smaller than $(T\log n)^{1/2}\epsilon$ for all $n \ge n_0$. This implies that $\max_{1\le i\le n} P\{Z_i > (T\log n)^{1/2}\epsilon\} \le n^{-2}$. Therefore, we have $\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}| = O_p\{(T/\log n)^{-1/2}\}$.

For the second result, by the first result and (A.13), we have $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = o_p\{(T/\log n)^{-1/2}\}$.

Step 4 (Conclusion). By Step 3, we may take $\delta_n = (T/\log n)^{-1/2}$ in Step 2. Thus, by Step 1, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) = \Gamma_n^{-1}\left\{-n^{-1}\sum_{i=1}^{n}\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\right\} + O_p\{(T/\log n)^{-3/4}\}. \tag{A.15}$$

The first term on the right side is $O_p\{(nT)^{-1/2}\}$. This shows that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\{(nT)^{-1/2} \vee (T/\log n)^{-3/4}\}$. If $n^2(\log n)^3/T \to 0$, then $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\{(nT)^{-1/2}\}$, and by the Lyapunov central limit theorem, we have $\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\{\mathbf{0}, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}\}$. □

Remark A.2. The reason why the order of the remainder term in (A.15) is $O_p\{(T/\log n)^{-3/4}\}$ and not $O_p(T^{-1})$ is that the exponent of δ_n inside the O_p terms on the right side of Eqs. (A.10) and (A.11) is 1/2 and not 1. Recall the definition of $g_{\alpha,\boldsymbol{\beta}}$ given in Step 2. Since $g_{\alpha,\boldsymbol{\beta}}$ is not differentiable with respect to $(\alpha, \boldsymbol{\beta})$, $E[g_{\alpha,\boldsymbol{\beta}}(\xi_{i1})^2]$ is bounded by const. $\times$ $(|\alpha| + \|\boldsymbol{\beta}\|)$ but not by const. $\times$ $(|\alpha|^2 + \|\boldsymbol{\beta}\|^2)$, which results in the exponent 1/2 of δ_n. Note that if $g_{\alpha,\boldsymbol{\beta}}$ were smooth in $(\alpha, \boldsymbol{\beta})$, we could use Taylor's theorem to bound $E[g_{\alpha,\boldsymbol{\beta}}(\xi_{i1})^2]$ by const. $\times$ $(|\alpha|^2 + \|\boldsymbol{\beta}\|^2)$. In that case, the exponent of δ_n would be 1, leading to the $O_p(T^{-1})$ rate of the remainder terms (we have ignored the $\log n$ term).

A.3. *Proof of Proposition 3.1*

The proof is basically similar to that of Kato and Galvao (2010, Theorem 3.2). However, as the present conditions are different from theirs, we give a proof of Proposition 3.1 for the sake of completeness.[14] Recall that under the present conditions, $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is weakly consistent. It suffices to show that uniformly over $1 \le i \le n$:

$$\frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it}) = f_i(0) + o_p(1),$$

$$\frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{x}_{it} = E[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}] + o_p(1),$$

[14] In fact, the condition on the bandwidth h_n is now weakened.

$$\frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{x}_{it}\boldsymbol{x}_{it}' = E[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}'] + o_p(1).$$

We only prove the first assertion because the proofs of the latter two assertions are analogous.

Without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. Put

$$f_i(\alpha, \boldsymbol{\beta}) := \frac{1}{T}\sum_{t=1}^{T} K_{h_n}(u_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta}).$$

We have to show that $\hat{f}_i(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) = f_i(0) + o_p(1)$ uniformly over $1 \le i \le n$. We first show that uniformly over $1 \le i \le n$,

$$\hat{f}_i(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) = E[\hat{f}_i(\alpha, \boldsymbol{\beta})]|_{\alpha=\hat{\alpha}_i, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + o_p(1).$$

To this end, it suffices to show that

$$\max_{1\le i\le n} \sup_{(\alpha,\boldsymbol{\beta})\in\mathbb{R}^{p+1}} |\hat{f}_i(\alpha, \boldsymbol{\beta}) - E[\hat{f}_i(\alpha, \boldsymbol{\beta})]| = o_p(1). \tag{A.16}$$

Define the class of functions $\mathcal{G}_{ni} := \{g_{\alpha,\boldsymbol{\beta},h_n} - E[g_{\alpha,\boldsymbol{\beta},h_n}(u_{i1}, \boldsymbol{x}_{i1})]: (\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}\}$ where $g_{\alpha,\boldsymbol{\beta},h}(u, \boldsymbol{x}) := K((u - \alpha - \boldsymbol{x}'\boldsymbol{\beta})/h)$. Put $Z_{ni} := \|\sum_{t=1}^{T} g(u_{it}, \boldsymbol{x}_{it})\|_{\mathcal{G}_i}$. By condition (C1), the class $\mathcal{G}_{ni}$ is uniformly bounded by some constant U (say) independent of i and n. By Bousquet's inequality (Proposition B.2), for all $s > 0$, with probability at least $1 - e^{-s^2}$,

$$Z_{ni} \le E[Z_{ni}] + s\sqrt{2(TC_f C_K h_n + 2UE[Z_{ni}])} + \frac{s^2 U}{3},$$

where we have used the fact that $E[g_{\alpha,\boldsymbol{\beta},h}(u_{i1}, \boldsymbol{x}_{i1})^2] = hE[\int K(u)^2 f_i(uh + \alpha + \boldsymbol{x}_{i1}'\boldsymbol{\beta}|\boldsymbol{x}_{i1})du] \le hC_f \int K(u)^2 du = hC_f C_K$ with $C_K := \int K(u)^2 du$. To estimate $E[Z_{ni}]$, we use Proposition B.1. The bounded variation of K on $\mathbb{R}$ guarantees that there exist positive constants $A \ge 3\sqrt{e}$ and $v \ge 1$ independent of i and n such that $N(\mathcal{G}_i, L_2(Q), U\epsilon) \le (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and every probability measure Q on $\mathbb{R}^{d+1}$ (cf. Nolan and Pollard (1987, Lemma 22)). Thus, by Proposition B.1, we have

$$E[Z_{ni}] \le \text{const.} \times \{\log n + (Th_n \log n)^{1/2}\},$$

where the constant is independent of i and n. Take $s = \sqrt{2\log n}$. Then, for each $1 \le i \le n$, with probability at least $1 - n^{-2}$, we have

$$Z_{ni} \le \text{const.} \times \{\log n + (Th_n \log n)^{1/2}\},$$

where the constant is independent of i and n. By the union bound and the present hypothesis that $\log n/(Th_n) \to 0$, we obtain (A.16).

The next step is to show that uniformly over $1 \le i \le n$,

$$E[\hat{f}_i(\alpha, \boldsymbol{\beta})]|_{\alpha=\hat{\alpha}_i, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = E[\hat{f}_i(0, \mathbf{0})] + o_p(1).$$

To see this, observe that

$$\begin{aligned}
&|E[\hat{f}_i(\alpha, \boldsymbol{\beta})] - E[\hat{f}_i(0, \mathbf{0})]| \\
&= \left|E\left[\int K(u)\{f_i(uh_n + \alpha + \boldsymbol{x}_{i1}'\boldsymbol{\beta}|\boldsymbol{x}_{i1}) - f_i(uh_n|\boldsymbol{x}_{i1})\}du\right]\right| \\
&\le L_f(|\alpha| + M\|\boldsymbol{\beta}\|).
\end{aligned}$$

Because of the weak consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$, we obtain the desired result.

The final step is to show that uniformly over $1 \le i \le n$,

$$E[\hat{f}_i(0, \mathbf{0})] = f_i(0) + o(1).$$

However, this can be derived from a standard calculation. The proof ends. □

A.4. Proof of Theorem 5.1

The proof is basically a modification of the proofs of Theorems 3.1 and 3.2 to the case where the data are dependent in the time dimension. To avoid duplication, we only point out the required modifications.

For the weak consistency, the only point that we need to change is the proof of (A.3). Instead of the Marcinkiewicz–Zygmund inequality, we now apply a Bernstein type inequality for β-mixing sequences (see Corollary C.1 below), by using Lemma C.1 to evaluate the variance term (see also the discussion following the lemma). Because of the exponential β-mixing property (condition (D1)), and the uniform boundedness of $\boldsymbol{x}_{it}$ (condition (B2)), taking $s = 2\log n$ and $q = [\sqrt{T}]$ in Corollary C.1, and using the fact that $(\log n)/\sqrt{T} \to 0$, we have that for any fixed $\epsilon > 0$, for large n,

$$\begin{aligned}
&\max_{1\le i\le n} P\left\{\sup_{(\alpha,\boldsymbol{\beta})\in B_i(\delta)} |\Delta_{ni}(\alpha, \boldsymbol{\beta}) - E[\Delta_{ni}(\alpha, \boldsymbol{\beta})]| > \epsilon\right\} \\
&\le \text{const.} \times \left(n^{-2} + \sqrt{T}Ba^{[\sqrt{T}]}\right).
\end{aligned}$$

Because $(\log n)/\sqrt{T} \to 0$, the right side is $o(n^{-1})$, which leads to the weak consistency.

For the expansion (3.2), we need some efforts. We will follow the notation in the proof of Theorem 3.2. First, the expansion (A.5) does not depend on the independence assumption and is valid under the present conditions. Second, instead of (A.10) and (A.11), we wish to prove for any $c \in (0, 1)$, provided that $|\log \delta_n| \asymp \log n$,

$$\begin{aligned}
&\left\|n^{-1}\sum_{i=1}^{n} \boldsymbol{\gamma}_i\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\}\right\| \\
&= O_p\{T^{-(1-c)}(\log n) \vee T^{-1/2}\delta_n^{1/2}(\log n)^{1/2}\},
\end{aligned} \tag{A.17}$$

$$\begin{aligned}
&\|\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\| \\
&= O_p\{T^{-(1-c)}(\log n) \vee T^{-1/2}\delta_n^{1/2}(\log n)^{1/2}\}.
\end{aligned} \tag{A.18}$$

As before, we only provide a proof for (A.17). Pick any $c \in (0, 1)$. As in the proof of Theorem 3.2, it suffices to show that

$$\begin{aligned}
&\max_{1\le i\le n} E\left[\left\|\sum_{t=1}^{T}\{g(\xi_{it}) - E[g(\xi_{i1})]\}\right\|_{\mathcal{G}_{\delta_n}}\right] \\
&= O_p\{T^{c}(\log n) \vee T^{1/2}\delta_n^{1/2}(\log n)^{1/2}\},
\end{aligned} \tag{A.19}$$

where $g_{\alpha,\boldsymbol{\beta}}(u, \boldsymbol{x}) := I(u \le \alpha + \boldsymbol{x}'\boldsymbol{\beta}) - I(u \le 0)$, $\mathcal{G}_\delta := \{g_{\alpha,\boldsymbol{\beta}}: |\alpha| \le \delta, \|\boldsymbol{\beta}\| \le \delta\}$ and $\xi_{it} := (u_{it}, \boldsymbol{x}_{it})$. Fix any $1 \le i \le n$. We apply Proposition C.1 to the class of functions $\tilde{\mathcal{G}}_{i,\delta_n} := \{g - E[g(\xi_{i1})]: g \in \mathcal{G}_{\delta_n}\}$. It is standard to see that $\tilde{\mathcal{G}}_{i,\delta_n}$ is uniformly bounded by $U = 2$ and there exist constants $A \ge 5e$ and $v \ge 1$ independent of i and n such that $N(\tilde{\mathcal{G}}_{i,\delta_n}, L_1(Q), 2\epsilon) \le (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and every probability measure Q on $\mathbb{R}^{p+1}$. Take $q = [T^c]$ and deduce from Lemma C.1 that $\sup_{g\in\tilde{\mathcal{G}}_{i,\delta_n}} \mathrm{Var}\{\sum_{t=1}^{q} g(\xi_{it})/\sqrt{q}\} \le \text{const.} \times \delta_n^{1/2}$ where the constant is independent of i and n (apply Lemma C.1 with $\delta = 1$; see also the discussion following the lemma). Since by condition (D1) $\max_{1\le i\le n} T\beta_i([T^c]) = o(1)$, we obtain (A.19).

We continue to prove the expansion (3.2). The conclusion of Step 3 in the proof of Theorem 3.2 follows from applying a Bernstein inequality and Talagrand's inequality for β-mixing sequences (Corollary C.1 and Proposition C.2) instead of Hoeffding's inequality and Talagrand's inequality for i.i.d. random variables used in the previous proof. Putting these together and taking c sufficiently small, we obtain the expansion (3.2).

Finally, we prove the asymptotic normality. Assume that $n^2(\log n)^3/T \to 0$. Then, we have the expansion

$$\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \{\Gamma^{-1} + o(1)\} \times \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\tau - I(u_{it} \le 0)\}(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)\right] + o_p\{(nT)^{-1/2}\}.$$

We wish to show a central limit theorem for the first term on the right side. Without loss of generality, we may assume that $\boldsymbol{x}_{it}$ is scalar. Put $z_{ni} := T^{-1/2}\sum_{t=1}^{T}\{\tau - I(u_{it} \le 0)\}(x_{it} - \gamma_i)$. Observe that $z_{n1}, \dots, z_{nn}$ are independent. Viewing that

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\tau - I(u_{it} \le 0)\}(x_{it} - \gamma_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_{ni},$$

we check the Lyapunov condition for the right sum. To this end, it suffices to show that $\sum_{i=1}^{n} E[|z_{ni}|^3] = o(n^{3/2})$. By conditions (B1) and (B2), $\{\tau - I(u_{it} \le 0)\}(x_{it} - \gamma_i)$ is uniformly bounded. By the exponential β-mixing property (condition (D1)) and Theorem 3 of Yoshihara (1978), we now deduce that $\max_{1 \le i \le n} E[|z_{ni}|^3] = O(1)$, which leads to that $\sum_{i=1}^{n} E[|z_{ni}|^3] = O(n) = o(n^{3/2})$. This completes the proof. □

Appendix B. Inequalities from empirical process theory: i.i.d. case

In this appendix, we introduce two inequalities from empirical process theory that were used in the proof of Theorem 3.2. Let $\xi_1, \dots, \xi_T$ be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$. The next proposition is a moment inequality for centered empirical processes, which is due to Proposition 2.2 of Gine and Guillou (2001). To avoid the measurability problem, we assume $\mathcal{F}$ to be a pointwise measurable class of functions, i.e., each element of $\mathcal{F}$ is measurable and there exists a countable subset $\mathcal{G} \subset \mathcal{F}$ such that for each $f \in \mathcal{F}$, there exists a sequence $\{g_m\} \subset \mathcal{G}$ with $g_m(\xi) \to f(\xi)$ for all $\xi \in S$. This condition is discussed in Section 2.3 of van der Vaart and Wellner (1996).

Proposition B.1. *Let $\mathcal{F}$ be a uniformly bounded, pointwise measurable class of functions on $(S, \mathcal{S})$ uniformly bounded by some constant U such that for some constants $A \ge 3\sqrt{e}$ and $v \ge 1$, $N(\mathcal{F}, L_2(Q), U\epsilon) \le (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and every probability measure Q on $(S, \mathcal{S})$. Moreover, suppose that $E[f(\xi_1)] = 0$ for all $f \in \mathcal{F}$. Let $\sigma^2 \ge \sup_{f \in \mathcal{F}} E[f^2(\xi_1)]$ be such that $0 < \sigma \le U$. Then, for all $T \ge 1$,*

$$E\left[\left\|\sum_{t=1}^{T} f(\xi_t)\right\|_{\mathcal{F}}\right] \le C\left[vU\log\frac{AU}{\sigma} + \sqrt{v}\sqrt{T}\sigma\sqrt{\log\frac{AU}{\sigma}}\right],$$

where C is a universal constant.

The next proposition is a Bernstein type inequality for centered empirical processes, which originates from Talagrand (1996). The current form of the inequality is due to Bousquet (2002).[15]

Proposition B.2. *Let $\mathcal{F}$ be a pointwise measurable class of functions on S uniformly bounded by some constant U. Moreover, suppose that $E[f(\xi_1)] = 0$ for all $f \in \mathcal{F}$. Let σ^2 be a positive constant such that $\sigma^2 \ge \sup_{f \in \mathcal{F}}[f^2(\xi_1)]$. Put $Z := \|\sum_{t=1}^{T} f(\xi_t)\|_{\mathcal{F}}$. Then, for all $s > 0$, we have*

$$P\left\{Z \ge E[Z] + s\sqrt{2(T\sigma^2 + 2UE[Z])} + \frac{s^2 U}{3}\right\} \le e^{-s^2}.$$

Appendix C. Some stochastic inequalities for β-mixing sequences

In this section, we introduce some stochastic inequalities for β-mixing sequences that we used in the proof of Theorem 5.1. Let $\{\xi_t, t \ge 1\}$ be a stationary process taking values in a measurable space $(S, \mathcal{S})$. We assume that S is a Polish space and $\mathcal{S}$ is its Borel σ-field. For a function f on S and a positive integer q, define $\sigma_q^2(f) := \mathrm{Var}\{f(\xi_1)\} + 2\sum_{j=1}^{q-1}(1 - j/q)\mathrm{Cov}\{f(\xi_1), f(\xi_{1+j})\}$, which is the variance of the sum $\sum_{t=1}^{q} f(\xi_t)/\sqrt{q}$. Let $\beta(j)$ denote the β-mixing coefficients of $\{\xi_t\}$. The next proposition is an extension of Proposition B.1 to β-mixing sequences.

Proposition C.1. *Let $\mathcal{F}$ be a pointwise measurable class of functions on S such that* (i) *for any $f \in \mathcal{F}$, $E[f(\xi_t)] = 0$;* (ii) *for any $f \in \mathcal{F}$, $\sup_{x \in S}|f(x)| \le U$;* (iii) *there exist constants $A \ge 5e$ and $v \ge 1$ such that $N(\mathcal{F}, L_1(Q), U\epsilon) \le (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and every probability measure Q on S. For any integer $q \in [1, T/2]$, let $\sigma_q^2 \ge \sup_{f \in \mathcal{F}} \sigma_q^2(f)$ be such that $0 < \sigma_q^2 \le 2qU^2$. Then, we have*

$$E\left[\left\|\sum_{t=1}^{T} f(\xi_t)\right\|_{\mathcal{F}}\right] \le C\left[qvU\log\frac{\sqrt{q}A'U}{\sigma_q} + \sqrt{v}\sqrt{T}\sigma_q\sqrt{\log\frac{\sqrt{q}A'U}{\sigma_q}}\right] + 2UT\beta(q), \qquad (C.1)$$

where C is a universal constant and $A' := \sqrt{2A}$.

Proof. The proof is based on Proposition 2 of Doukhan et al. (1995), which is deduced from Berbee's (1979) coupling lemma, and the proof of Proposition 2.1 in Gine and Guillou (2001). Use Proposition 2 of Doukhan et al. (1995) to construct a sequence $\{\tilde{\xi}_t\}_{t \ge 1}$ such that (a) $\tilde{\Xi}_k := (\tilde{\xi}_{1+(k-1)q}, \dots, \tilde{\xi}_{kq})$ has the same distribution as $\Xi_k := (\xi_{1+(k-1)q}, \dots, \xi_{kq})$; (b) $P(\Xi_k \ne \tilde{\Xi}_k) \le \beta(q)$; (c) $\{\tilde{\Xi}_{2k}, k \ge 1\}$ are independent and so are $\{\tilde{\Xi}_{2k-1}, k \ge 1\}$. Put $r := [T/(2q)]$. With a slight abuse of notation, for a function f on S, we write $f(\Xi_k) = \sum_{t \in T_k} f(\xi_t)$, where $T_k := \{1 + (k-1)q, \dots, kq\}$. Then, we have

$$\begin{aligned} E\left[\left\|\sum_{t=1}^{T} f(\xi_t)\right\|_{\mathcal{F}}\right] &\le E\left[\left\|\sum_{t=1}^{T} f(\tilde{\xi}_t)\right\|_{\mathcal{F}}\right] + 2UE\left[\sum_{t=1}^{T} I(\xi_t \ne \tilde{\xi}_t)\right] \\ &\le E\left[\left\|\sum_{t=1}^{T} f(\tilde{\xi}_t)\right\|_{\mathcal{F}}\right] + 2UT\beta(q) \\ &\le 2E\left[\left\|\sum_{k=1}^{r} f(\tilde{\Xi}_{2k})\right\|_{\mathcal{F}}\right] + (T - 2qr)U + 2UT\beta(q), \end{aligned} \qquad (C.2)$$

[15] Talagrand's (1996) Theorem 1.4 assumes $\mathcal{F}$ to be a countable class. Clearly, this condition can be weakened to the case where $\mathcal{F}$ is pointwise measurable.

where the second inequality is due to the fact that $\{\tilde{\Xi}_{2k-1}, 1 \le k \le r\}$ has the same distribution as $\{\tilde{\Xi}_{2k}, 1 \le k \le r\}$. Let $\epsilon_1, \dots, \epsilon_r$ be i.i.d. random variables with $P(\epsilon_k = \pm 1) = 1/2$ independent of $\{\tilde{\Xi}_{2k}, 1 \le k \le r\}$. Recall that $\tilde{\Xi}_{2k}, 1 \le k \le r$ are i.i.d. By Lemma 2.3.1 of van der Vaart and Wellner (1996), we have

$$E\left[\left\|\sum_{k=1}^r f(\tilde{\Xi}_{2k})\right\|_{\mathcal{F}}\right] \le 2E\left[\left\|\sum_{k=1}^r \epsilon_k f(\tilde{\Xi}_{2k})\right\|_{\mathcal{F}}\right] = 2qUE\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})\right\|_{\mathcal{H}}\right], \qquad (C.3)$$

where $\mathcal{H} := \{\varphi(\xi_1, \dots, \xi_q) = \sum_{t=1}^q f(\xi_t)/(qU) : f \in \mathcal{F}\}$. We shall bound the right side of (C.3). Without loss of generality, we may assume that $0 \in \mathcal{H}$. By Hoeffding's inequality, given $\tilde{\Xi}_{2k}, 1 \le k \le r$, the process $\varphi \mapsto \sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})/\sqrt{r}$ is sub-Gaussian for the $L_2(\tilde{Q}_r)$ norm, where $\tilde{Q}_r$ is the empirical distribution on S^q that assigns probability $1/r$ to each even block $\tilde{\Xi}_{2k}, 1 \le k \le r$. Thus, by Corollary 2.2.8 of van der Vaart and Wellner (1996), we have

$$E_\epsilon\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})/\sqrt{r}\right\|_{\mathcal{H}}\right] \le C\int_0^{\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})/r\right\|_{\mathcal{H}}^{1/2}} \sqrt{\log N(\mathcal{H}, L_2(\tilde{Q}_r), \tau)}\,d\tau,$$

where E_ϵ stands for the expectation with respect to ϵ_k's and C is a universal constant. Let $\tilde{P}_{qr}$ denote the empirical distribution on S that assigns probability $1/(qr)$ to each $\tilde{\xi}_t, t \in \cup_{k=1}^r T_{2k}$. Since for $\varphi_i(\xi_1, \dots, \xi_q) = \sum_{t=1}^q f_i(\xi_t)/(qU), f_i \in \mathcal{F}, i = 1, 2,$

$$\begin{aligned}\frac{1}{r}\sum_{k=1}^r \{\varphi_1(\tilde{\Xi}_{2k}) - \varphi_2(\tilde{\Xi}_{2k})\}^2 &= \frac{1}{q^2 r U^2}\sum_{k=1}^r \{f_1(\tilde{\Xi}_{2k}) - f_2(\tilde{\Xi}_{2k})\}^2 \\ &\le \frac{2}{qrU}\sum_{k=1}^r |f_1(\tilde{\Xi}_{2k}) - f_2(\tilde{\Xi}_{2k})| \\ &\le \frac{2}{U}\|f_1 - f_2\|_{L_1(\tilde{P}_{qr})},\end{aligned}$$

we have $N(\mathcal{H}, L_2(\tilde{Q}_r), \tau) \le N(\mathcal{F}, L_1(\tilde{P}_{qr}), U\tau^2/2) \le (2A/\tau^2)^v$. Thus,

$$\begin{aligned}E_\epsilon\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})/\sqrt{r}\right\|_{\mathcal{H}}\right] &\le C\sqrt{2v}\int_0^{\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})/r\right\|_{\mathcal{H}}^{1/2}} \sqrt{\log\left(\sqrt{2A}/\tau\right)}\,d\tau \\ &= 2C\sqrt{Av}\int_{\sqrt{2A}/\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})/r\right\|_{\mathcal{H}}^{1/2}}^{\infty} \frac{\sqrt{\log\tau}}{\tau^2}\,d\tau.\end{aligned}$$

Integration by parts gives

$$\begin{aligned}\int_a^\infty \frac{\sqrt{\log\tau}}{\tau^2}d\tau &= \left[-\frac{\sqrt{\log\tau}}{\tau}\right]_a^\infty + \frac{1}{2}\int_a^\infty \frac{1}{\tau^2\sqrt{\log\tau}}d\tau \\ &\le \frac{\sqrt{\log a}}{a} + \frac{1}{2}\int_a^\infty \frac{\sqrt{\log\tau}}{\tau^2}d\tau, \quad a \ge e,\end{aligned}$$

from which we have

$$\int_a^\infty \frac{\sqrt{\log\tau}}{\tau^2}d\tau \le \frac{2\sqrt{\log a}}{a}, \quad a \ge e.$$

Therefore, we have

$$\begin{aligned}E\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})/\sqrt{r}\right\|_{\mathcal{H}}\right] &\le 2C\sqrt{v}E\left[\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})/r\right\|_{\mathcal{H}}^{1/2}\sqrt{\log\frac{2A}{\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})/r\right\|_{\mathcal{H}}}}\right] \\ &\le 2C\sqrt{v}\sqrt{E\left[\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})/r\right\|_{\mathcal{H}}\right]\log\frac{2A}{E\left[\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})/r\right\|_{\mathcal{H}}\right]}},\end{aligned}$$

where the second inequality is due to Hölder's inequality, the concavity of the map $x \mapsto x\log(a/x)$ and Jensen's inequality.

Now, by Corollary 3.4 of Talagrand (1994),

$$E\left[\left\|\sum_{k=1}^r \varphi^2(\tilde{\Xi}_{2k})\right\|_{\mathcal{H}}\right] \le \frac{r\sigma_q^2}{qU^2} + 8E\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})\right\|_{\mathcal{H}}\right],$$

and the right side is bounded by $10r$ because $\sigma_q^2 \le 2qU^2$. Since the map $x \mapsto x\log(a/x)$ is non-decreasing for $0 \le x \le a/e$ and $A \ge 5e$, we have

$$E\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})/\sqrt{r}\right\|_{\mathcal{H}}\right] \le 2C\sqrt{v}\sqrt{\left(\frac{\sigma_q^2}{qU^2} + \frac{8}{r}E\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})\right\|_{\mathcal{H}}\right]\right)\log\frac{2qAU^2}{\sigma_q^2}}.$$

Put

$$Z := E\left[\left\|\sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_{2k})\right\|_{\mathcal{H}}\right].$$

Then, Z satisfies

$$Z^2 \le C\frac{vr\sigma_q^2}{qU^2}\log\frac{\sqrt{q}A'U}{\sigma_q} + 8CvZ\log\frac{\sqrt{q}A'U}{\sigma_q},$$

where C is another universal constant and $A' := \sqrt{2A}$. This gives

$$\begin{aligned}Z &\le 4Cv\log\frac{\sqrt{q}A'U}{\sigma_q} + \sqrt{16C^2v^2\left(\log\frac{\sqrt{q}A'U}{\sigma_q}\right)^2 + C\frac{vr\sigma_q^2}{qU^2}\log\frac{\sqrt{q}A'U}{\sigma_q}} \\ &\le 8Cv\log\frac{\sqrt{q}A'U}{\sigma_q} + \sqrt{C}\frac{\sqrt{v}\sqrt{r}\sigma_q}{\sqrt{q}U}\sqrt{\log\frac{\sqrt{q}A'U}{\sigma_q}} \\ &\le C'\left[v\log\frac{\sqrt{q}A'U}{\sigma_q} + \frac{\sqrt{v}\sqrt{r}\sigma_q}{\sqrt{q}U}\sqrt{\log\frac{\sqrt{q}A'U}{\sigma_q}}\right], \qquad (C.4)\end{aligned}$$

where the second inequality is due to $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for $a, b > 0$, and C' is another universal constant. Combining (C.2)–(C.4) yields the desired inequality. Note that $(T - 2qr)U$ is absorbed into $CqvU\log(\sqrt{q}A'U/\sigma_q)$ since $T - 2qr \le 2q$ and $\sqrt{q}A'U/\sigma_q > e$ under our assumption. □

Proposition C.2 and Corollary C.1 are due to Kato and Galvao (2010).

Proposition C.2 (*Talagrand's Inequality for β-Mixing Sequences*). *Suppose that the conditions of Proposition C.1 are satisfied. Assume that*

$$T\sigma_q^2 \geq q^2 \upsilon U^2 \log \frac{\sqrt{q}A'U}{\sigma_q},$$

where $A' := \sqrt{2A}$. Then, for all $s > 0$, we have

$$P\left\{\left\|\sum_{t=1}^{T} f(\xi_t)\right\|_{\mathcal{F}} \geq C\sqrt{\upsilon}\sqrt{T}\sigma_q\sqrt{\log \frac{\sqrt{q}A'U}{\sigma_q}} + C\sigma_q\sqrt{sT} + sqCU\right\} \leq 2e^{-s} + 2r\beta(q),$$

where $r := [T/(2q)]$ and C is a universal constant.

Corollary C.1 (*Bernstein's Inequality for β-Mixing Sequences*). *Let f be a function on S such that $\sup_{x\in S}|f(x)| \leq U$ and $E[f(\xi_1)] = 0$. Pick any $q \in [1, T/2]$. Then, for all $s > 0$, we have*

$$P\left\{\left|\sum_{t=1}^{T} f(\xi_t)\right| \geq C\left\{\sqrt{(s\vee 1)T}\sigma_q(f) + sqU\right\}\right\} \leq 2e^{-s} + 2r\beta(q),$$

where $r := [T/(2q)]$ and C is a universal constant.

In applying those inequalities, the evaluation of the variance term $\sigma_q^2(f)$ is essential. For β-mixing processes, Yoshihara's (1976) Lemma 1 is particularly useful for that purpose. Since it is repeatedly used in the proofs of the theorems above, we describe a special case of that lemma.

Lemma C.1 (Yoshihara (1976)). *Let j be a fixed positive integer. Let f and g be functions on S such that $E[f(\xi_1)] = E[g(\xi_{1+j})] = 0$, and for some positive constants δ and M,*

$$E[|f(\xi_1)|^{1+\delta}]E[|g(\xi_{1+j})|^{1+\delta}] \leq M, \qquad E[|f(\xi_1)g(\xi_{1+j})|^{1+\delta}] \leq M. \tag{C.5}$$

Then, we have

$$|\mathrm{Cov}(f(\xi_1), g(\xi_{1+j}))| \leq 4M^{1/(1+\delta)}\beta(j)^{\delta/(1+\delta)}.$$

A direct consequence of Lemma C.1 is that if there exist positive constants δ and M such that (C.5) holds for any positive integer j and $\sum_{j=1}^{\infty}\beta(j)^{1/(1+\delta)} < \infty$, then the infinite sum $\sum_{j=1}^{\infty}\mathrm{Cov}\{f(\xi_1), g(\xi_{1+j})\}$ is absolutely convergent, and in particular, for any positive integer q,

$$\mathrm{Var}\left\{\sum_{t=1}^{q} f(\xi_t)/\sqrt{q}\right\} \leq 4M^{1/(1+\delta)}\left\{1 + 2\sum_{j=1}^{\infty}\beta(j)^{\delta/(1+\delta)}\right\}.$$

If $\beta(j)$ decays exponentially fast as $j \to \infty$, i.e., for some constants $a \in (0, 1)$ and $B > 0$, $\beta(j) \leq Ba^j$, then $\sum_{j=1}^{\infty}\beta(j)^{\delta/(1+\delta)} < \infty$ for any $\delta > 0$.

References

Abrevaya, J., Dahl, C.M., 2008. The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. Journal of Business and Economic Statistics 26, 379–397.

Angrist, J., Chernozhukov, V., Fernandez-Val, I., 2006. Quantile regression under misspecification, with an application to the US wage structure. Econometrica 74, 539–563.

Arcones, M.A., 1998. Second order representations of the least absolute deviation regression estimator. Annals of the Institute of Statistical Mathematics 50, 87–117.

Berbee, H.C.P., 1979. Random Walks with Stationary Increments and Renewal Theory, Mathematical Centre Tracts 112, Mathematisch Centrum, Amsterdam.

Bousquet, O., 2002. A bennett concentration inequality and its application to suprema of empirical processes. Comptes Rendus Mathematique 334, 495–500.

Canay, I.A., 2008. A Note on Quantile Regression for Panel Data Models. Mimeo, Northwestern University.

Chernozhukov, V., Fernandez-Val, I., Newey, W., 2009, Quantile and average effects in nonseparable panel models, CEMMAP Working Paper CWP29/09.

Chow, Y.S., Teicher, H., 1997. Probability Theory, third ed. Springer-Verlag, New York.

Doukhan, P., Massart, P., Rio, E., 1995. Invariance principles for absolutely regular empirical processes. Annales de l'Institut Henri Poincare 31, 393–427.

Fernandez-Val, I., 2005. Bias Correction in Panel Data Models with Individual Specific Parameters. Mimeo.

Gine, E., Guillou, A., 2001. On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. Annales de l'Institut Henri Poincare 37, 503–522.

Graham, B.S., Hahn, J., Powell, J.L., 2009. the incidental parameter problem in a non-differentiable panel data model. Economics Letters 105, 181–182.

Gutenbrunner, C., Jureckova, J., 1992. Regression rank scores and regression quantiles. Annals of Statistics 20, 305–330.

Hahn, J., Kuersteiner, G.M., 2004. Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects. Mimeo.

Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72, 1295–1319.

He, X., Shao, Q.-M., 2000. On parameters of increasing dimensions. Journal of Multivariate Analysis 73, 120–135.

Imbens, G.W., Newey, W.K., 2009. Identification and estimation of triangular simultaneous equation models without additivity. Econometrica 77, 1481–1512.

Kato, K., Galvao, A.F., 2010. Smoothed Quantile Regression for Panel Data. Mimeo.

Knight, K., 1998. Limiting distributions for L_1 regression estimators under general conditions. Annals of Statistics 26, 755–770.

Koenker, R., 1984. A note on L-estimates for linear models. Statistics and Probability Letters 2, 323–325.

Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.

Koenker, R., Bassett, G.W., 1978. Regression quantiles. Econometrica 46, 33–49.

Koenker, R., Hallock, K., 2000, Quantile Regression: An Introduction, Manuscript, University of Illinois at Urbana-Champaign.

Lamarche, C., 2010. Robust penalized quantile regression estimation for panel data. Journal of Econometrics 157, 396–408.

Lancaster, T., 2000. The incidental parameter problem since 1948. Journal of Econometrics 95, 391–413.

Li, H., Lindsay, B.G., Waterman, R.P., 2003. Efficiency of projected score methods in rectangular array asymptotics. Journal of the Royal Statistical Society, Series B 65, 191–208.

Neyman, J., Scott, E.L., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.

Nolan, D., Pollard, D., 1987. U-processes: rates of convergence. Annals of Statistics 15, 780–799.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.

Rosen, A., 2009, Set identification via quantile restrictions in short panels, CEMMAP Working Paper CWP26/09.

Talagrand, M., 1994. Sharper bounds for Gaussian and empirical processes. Annals of Probability 22, 28–76.

Talagrand, M., 1996. New concentration inequalities in product spaces. Inventiones Mathematicae 126, 505–563.

van der Vaart, A., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag, New York, New York.

Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 35, 237–252.

Yoshihara, K., 1978. Moment inequalities for mixing sequences. Kodai Mathematical Journal 1, 316–328.