Smoothed quantile regression for panel data*

Kengo Kato[†] Antonio F. Galvao, Jr.[‡]

June 11, 2010

Abstract

This paper studies fixed effects estimation of quantile regression (QR) models with panel data. Previous studies show that there are two important difficulties with the standard QR estimation. First, the estimator can be biased because of the well-known incidental parameters problem. Secondly, the non-smoothness of the objective function significantly complicates the asymptotic analysis of the estimator especially in the panel data case. We overcome the latter problem by smoothing the objective function. Under an asymptotic framework where both the numbers of individuals and time periods grow at the same rate, we show that the fixed effects estimator for the smoothed objective function has a limiting normal distribution with a bias in the mean, giving the analytic form of the asymptotic bias. We propose a simple one-step bias correction to the fixed effects estimator based on the analytic bias formula obtained from our asymptotic analysis. We illustrate the effect of the bias correction to the estimator through simulations.

Key words: bias correction, incidental parameters problem, panel data, quantile regression, smoothing.

JEL classification codes: C13, C23.

*Preliminary draft.

[†]Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. E:mail: `kkato@hiroshima-u.ac.jp`

[‡]Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201. E-mail: `agalvao@uwm.edu`

1 Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression (QR) has attracted considerable interest in statistics and econometrics. It offers an easy-to-implement method to estimate conditional quantiles, and is known as a more flexible tool to capture the effect of explanatory variables to the response than mean regression. The theoretical properties of the QR method are well established for cross sectional models; see Koenker (2005) for references and discussion.

In contrast to mean regression, however, little is known about the ability of the QR method to analyze panel (longitudinal) data models in which individual effects are included to capture unobserved heterogeneity. An approach in this direction is to treat each individual effect as a parameter to be estimated and apply the standard QR estimation to the individual and common parameters all together (Koenker, 2004).[1] Unfortunately, the resulting estimator of the common parameters in general will be inconsistent when the number of individuals n goes to infinity while the number of time periods T is fixed. This is a version of incidental parameters problems (Neyman and Scott, 1948) which have been extensively studied in the recent econometrics literature (see Lancaster, 2000, for a review). Arellano and Bonhomme (2009, p.490) stated that the incidental parameters problem is "one of the main challenges in modern econometrics".

A recent attempt to cope with the incidental parameters problem is to introduce an asymptotic framework that n and T jointly go to infinity. Li, Lindsay, and Waterman (2003) considered the maximum likelihood (ML) estimation for smooth likelihood functions and showed that the ML estimator (MLE) of the common parameters has an order $O(T^{-1})$ bias, so its limiting normal distribution has a bias in the mean (even) when n and T grow at the same rate. An important implication of their work is that properly bias corrected estimators enjoy mean-zero asymptotic normality for T with a slower rate than n. Since then, several bias correction methods for the ML estimation (or more generally, M-estimation) have been proposed in the literature; see Hahn and Newey (2004), Woutersen (2002), Arellano and Hahn (2005), Bester and Hansen (2009), Arellano and Bonhomme (2009) and Dhaene and Jochmans (2009), to name only a few.

A distinctive feature of QR is that the objective function, often referred to as the check function, is not differentiable. This means that the asymptotic analysis of the previous nonlinear panel data literature is not directly applicable to the QR

[1] Koenker (2004) indeed proposed a penalized estimation method where the individual parameters are subject to the ℓ_1 penalty. Other approaches are found in Abrevaya and Dahl (2008), Canay (2008) and Rosen (2009).

case since it substantially depends on the smoothness of objective functions. Kato, Galvao, and Montes-Rojas (2010) formally established the asymptotic properties of the standard QR estimator of the common parameters. However, they required the restrictive condition that T grows faster than n^2 to show asymptotic normality of the estimator, and did not succeed in deriving the bias. The difficulty to handle the QR estimator in panel models is partly explained by the fact that the bias expression in Hahn and Newey (2004) is related to the second order bias of the MLE for the individual parameters while the second order behavior of QR estimators is non-standard and rather complicated (Arcones, 1998; Knight, 1998). It is worthwhile to remark that, to our knowledge, there is no paper that formally derives the bias for the standard QR estimator under the large n and T asymptotics, nor rigorously studies the bias correction for panel QR models.

This paper overcomes the above problem by smoothing the objective function. The idea of smoothing non-differentiable objective functions owes to Amemiya (1982) and Horowitz (1992, 1998). We refer to the resulting estimator as the fixed effects smoothed quantile regression (FE-SQR) estimator. We show that under suitable regularity conditions, the FE-SQR estimator has an order $O(T^{-1})$ bias and hence has a limiting normal distribution with a bias in the mean (even) when n and T grow at the same rate. As naturally expected, the bias depends on the conditional and unconditional densities and their first derivatives. We propose a simple one-step bias correction based on the analytic form of the asymptotic bias. We also examine the half-panel jackknife method originally proposed by Dhaene and Jochmans (2009) to the FE-SQR estimator. We theoretically show that the both methods eliminate the bias of the limiting distribution. It is important to note that these asymptotic results are not derived from the method of Li, Lindsay, and Waterman (2003) since the bandwidth tending to zero as the sample size increases is involved in the objective function and the standard stochastic expansion does not apply to such a case (see Horowitz, 1998). In this case, we have to control the smoothing effect and at the same time handle the problem of diverging number of parameters, which we believe is challenging from a technical point of view.

We conduct a small Monte Carlo study to illustrate the effect of the bias correction to the FE-SQR estimator. We examine the cases where $n \in \{100, 200\}$ and $T \in \{8, 12, 16, 20\}$. The results show that, on the one hand, as expected, the standard QR and the FE-SQR estimators are moderately biased except for some special cases. In addition, the one-step bias correction is able to substantially reduce the bias in many cases, although it slightly increases the variability in the small sample partly because of the nonparametric estimation of the bias term.

The organization of this paper is as follows. In Section 2, we give a QR model with individual effects and formally define the FE-SQR estimator. In Section 3, we present the theorems about the limiting distributions of the FE-SQR estimator and the bias corrected one when n and T grow at the same rate. In Section 4, we report a Monte Carlo study to assess the finite sample performance of the estimator and the bias correction. In Section 5, we leave some concluding remarks. We relegate the proofs of the theorems to the Appendix.

2 Model and estimation method

For each fixed $\tau \in (0,1)$, we consider a QR model with individual effects

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

where y_{it} is a response variable, $\boldsymbol{x}_{it}$ is a p dimensional vector of explanatory variables and $Q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_i^*)$ is the conditional τ-quantile of y_{it} given $\boldsymbol{x}_{it}$ and α_i^*. In model (2.1), each α_i^* represents an individual heterogeneity and is called an individual effect. In general, each α_i^* and $\boldsymbol{\beta}_0$ can depend on τ, but we assume τ to be fixed throughout the paper and suppress such a dependence for notational simplicity. The number of individuals is denoted by n and the number of time periods is denoted by $T = T_n$ which may depends on n. Usually, we omit the subscript n of T_n. It should be noted that although the model (2.1) looks “linear” at a first sight, the quantile is a nonlinear functional of the random variable, so the conventional differencing out strategy is not appropriate for the QR case.

We consider the fixed effects estimation of $\boldsymbol{\beta}_0$, which is implemented by treating each individual effect as a parameter to be estimated. The standard QR estimation solves

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta}) \{ \tau - I(y_{it} \le \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}) \} \right], \tag{2.2}$$

where $I(\cdot)$ is the indicator function, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)' \in \mathbb{R}^n$ and $\boldsymbol{\beta} \in \mathbb{R}^p$. Note that $\boldsymbol{\alpha}$ implicitly depends on n. As in general nonlinear panel models, except for some special cases, the estimator of $\boldsymbol{\beta}_0$ defined by (2.2) is inconsistent as $n \to \infty$ while T is fixed. For completeness, we give in Appendix A a simple example in which the standard QR estimator of the common parameter is actually inconsistent as $n \to \infty$ while T is fixed, since we were not aware of such an example in published works. It is interesting to note that in the example described in Appendix A the bias can be arbitrarily large by letting τ close to zero or one.

A distinctive feature of the standard QR estimation is that the objective function is not differentiable. Therefore, the basic smoothness assumption imposed in the recent nonlinear panel data literature is not satisfied for the standard QR estimator. Kato, Galvao, and Montes-Rojas (2010) rigorously investigated the asymptotic properties of the standard QR estimator, say $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$, defined by (2.2) when n and T jointly go to infinity. The essential content of their theoretical results is that while $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$ is consistent under mild regularity conditions, justifying (mean-zero) asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$ requires the restrictive condition that T grows faster than n^2.[2] They remarked that the non-smoothness of the objective function (more precisely, the non-smoothness of the scores) significantly complicates the asymptotic analysis of $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$ when n and T jointly go to infinity. In particular, to our knowledge, whether the asymptotic results analogous to those of Li, Lindsay, and Waterman (2003) or Hahn and Newey (2004) hold for $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$ is not known.

In this paper, instead of the standard QR estimator, we study the asymptotic properties of the estimator defined by a minimizer of a smoothed version of the QR objective function. Smoothing the QR objective function is employed in Horowitz (1998) to study the bootstrap refinement for inference in conditional quantile models.[3] The basic insight of Horowitz (1998) is to smooth over $I(y_{it} \leq \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})$ by using a kernel function. Let $K(\cdot)$ be a kernel function and $G(\cdot)$ be the survival function of $K(\cdot)$, i.e.,

$$\int_{-\infty}^{\infty} K(u)du = 1,\ G(u) := \int_{u}^{\infty} K(v)dv.$$

We do not require $K(\cdot)$ to be nonnegative, and will state some requirements of $K(\cdot)$ in Section 3. Let $\{h_n\}$ be a sequence of positive numbers (bandwidths) such that $h_n \to 0$ as $n \to \infty$ and write $G_{h_n}(\cdot) = G(\cdot/h_n)$. Note that $G_{h_n}(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta})$ is a smoothed counterpart of $I(y_{it} \leq \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})$. Then, we consider the estimator

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) := \arg \min_{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}^n \times \mathcal{B}} \left[\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta}) \{ \tau - G_{h_n}(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta}) \} \right], \tag{2.3}$$

where $\mathcal{A}$ is a compact subset of $\mathbb{R}$, $\mathcal{A}^n$ is the product of n copies of $\mathcal{A}$ and $\mathcal{B}$ is a compact subset of $\mathbb{R}^p$. The compactness of $\mathcal{A}^n \times \mathcal{B}$ is required to ensure the existence of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. We call $\hat{\boldsymbol{\beta}}$ the fixed effects smoothed quantile regression (FE-SQR) estimator of $\boldsymbol{\beta}_0$.

Put $\tilde{K}(u) := uK(u)$ and $\tilde{K}_{h_n}(u) := h_n^{-1}\tilde{K}(u/h_n)$. We use the notation $\partial_{\alpha_i} = \partial/\partial\alpha_i$

[2] Kato, Galvao, and Montes-Rojas (2010), however, did not show that this condition is necessary for (mean-zero) asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$. Whether one can substantially improve this condition is still an open question.

[3] A motivation to handle the smoothed QR estimator in Horowitz (1998) is that higher order asymptotic theory of the unsmoothed QR estimator is non-standard, which significantly complicates the study of the bootstrap refinement for inference based on it (see also Horowitz, 2001, Section 4.3).

and $\partial_{\boldsymbol{\beta}} = \partial/\partial\boldsymbol{\beta}$. Since the objective function in (2.3) is smooth with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, if $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is an interior point of $\mathcal{A}^n \times \mathcal{B}$, it satisfies

$$\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) = 0, \ i = 1, \ldots, n; \ \mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \mathbf{0}, \tag{2.4}$$

where

$$\begin{aligned} \mathbb{H}_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}) &:= T^{-1} \textstyle\sum_{t=1}^T \{\tau - G_{h_n}(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta})\} \\ &\quad + h_n T^{-1} \textstyle\sum_{t=1}^T \tilde{K}_{h_n}(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta}), \\ \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &:= (nT)^{-1} \textstyle\sum_{i=1}^n \sum_{t=1}^T \{\tau - G_{h_n}(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta})\}\boldsymbol{x}_{it} \\ &\quad + h_n (nT)^{-1} \textstyle\sum_{i=1}^n \sum_{t=1}^T \tilde{K}_{h_n}(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta})\boldsymbol{x}_{it}. \end{aligned}$$

As we will see in the proof of Theorem 3.1, the terms concerning $\tilde{K}(\cdot)$ do not affect the first order asymptotic distribution of $\hat{\boldsymbol{\beta}}$. This can be understood by regarding $\tilde{K}(\cdot)$ as a "kernel", though the integral of $\tilde{K}(\cdot)$ over the real line is not one.

For a later purpose, we refer to the partial derivative with respect to α_i as the α_i-derivative and the partial derivative with respect to $\boldsymbol{\beta}$ as the $\boldsymbol{\beta}$-derivative. $n^{-1}\mathbb{H}_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta})$ and $\mathbb{H}_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are the α_i-derivative and the $\boldsymbol{\beta}$-derivative of one minus the objective function in (2.3), respectively. From an analogy to the ML estimation, we call $\mathbb{H}_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta})$ the α_i-score and $\mathbb{H}_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ the $\boldsymbol{\beta}$-score.

3 Asymptotic theory

In this section, we investigate the asymptotic properties of the FE-SQR estimator. In the asymptotic analysis, as in Hahn and Newey (2004) and Fernandez-Val (2005), we pick a single realization $(\alpha_{10}, \alpha_{20}, \ldots)$ of $(\alpha_1^*, \alpha_2^*, \ldots)$ and treat each α_{i0} as a fixed true parameter. In what follows, suppose that every argument is conditional on $\alpha_i^* = \alpha_{i0}$ for each individual i. Put $\boldsymbol{\alpha}_0 := (\alpha_{10}, \ldots, \alpha_{n0})'$.

3.1 Assumptions

This section provides conditions under which the FE-SQR estimator has a limiting normal distribution with a bias in the mean when n and T grow at the same rate. The conditions below are stronger than needed if the main concern is to prove consistency only or asymptotic normality of the estimator when $n/T \to 0$. However, the conditions are, to some extent, standard in the literature.

(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) There exists a constant M such that $\sup_{i\geq 1}\|\boldsymbol{x}_{it}\| \leq M$ (a.s.).

(A3) $\mathcal{A} \subset \mathbb{R}$ and $\mathcal{B} \subset \mathbb{R}^p$ are compact. For each i, α_{i0} is an interior point of $\mathcal{A}$, and $\boldsymbol{\beta}_0$ is an interior point of $\mathcal{B}$.

The distribution of $(y_{it}, \boldsymbol{x}_{it})$ is allowed to depend on i. Define $u_{it} := y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0$. 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}. For two sequences of positive numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \asymp b_n$ when there exists a positive constant C such that $C^{-1}b_n \leq a_n \leq Cb_n$ for every n.

(A4) The minimum eigenvalues of $\mathrm{E}[f_i(0|\boldsymbol{x}_{i1})(1, \boldsymbol{x}_{i1}')'(1, \boldsymbol{x}_{i1}')]$ are bounded away from zero uniformly over $i \geq 1$.

(A5) (a) For each i, $f_i(u|\boldsymbol{x})$ is $(r-1)$ times continuously differentiable with respect to u for each $\boldsymbol{x}$, where $r \geq 4$ is an even integer. Let $f_i^{(j)}(u|\boldsymbol{x}) := (\partial/\partial u)^j f_i(u|\boldsymbol{x})$ for $j = 0, 1, \ldots, r-1$, where $f_i^{(0)}(u|\boldsymbol{x})$ stands for $f_i(u|\boldsymbol{x})$. (b) There exists a constant A_0 such that $|f_i^{(j)}(u|\boldsymbol{x})| \leq A_0$ for $j = 0, 1, \ldots, r-1$ uniformly over both $(u, \boldsymbol{x})$ and i. (c) $\inf_{i\geq 1} f_i(0) > 0$.

(A6) (a) $K(\cdot)$ is symmetric about the origin, of bounded support and three times continuously differentiable. (b) $K(\cdot)$ is an r-th order kernel, that is,

$$\int_{-\infty}^{\infty} K(u)du = 1, \ \int_{-\infty}^{\infty} u^j K(u)du = 0, \ j = 1, \ldots, r-1, \ \int_{-\infty}^{\infty} u^r K(u)du \neq 0,$$

where r is given in condition (A4).

(A7) $h_n \asymp T^{-a}$, where $1/r < a < 1/3$.

Put $s_i := 1/f_i(0), \boldsymbol{\gamma}_i := s_i\mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}]$ and $\boldsymbol{\nu}_i := f_i^{(1)}(0)\boldsymbol{\gamma}_i - \mathrm{E}[f_i^{(1)}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}]$.

(A8) (a) $\Gamma_n := n^{-1}\sum_{i=1}^n \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}(\boldsymbol{x}_{i1}' - \boldsymbol{\gamma}_i')]$ is nonsingular for each n, and the limit $\Gamma := \lim_{n\to\infty}\Gamma_n$ exists and is nonsingular. (b) The limit

$$V := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{E}[(\boldsymbol{x}_{i1} - \boldsymbol{\gamma}_i)(\boldsymbol{x}_{i1} - \boldsymbol{\gamma}_i)']$$

exists and is nonsingular.

(A9) The limit $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n s_i^2\boldsymbol{\nu}_i$ exists.

Condition (A1) is the same as Condition 1 (i) in Fernandez-Val (2005). We exclude temporal dependence to ease the technical treatment.[4] The condition that $\boldsymbol{x}_{it}$ is uniformly bounded can be removed at the expense of the lengthier proof. Condition (A3) is standard in asymptotic theory. As Kato, Galvao, and Montes-Rojas (2010) noted, condition (A4) guarantees identification of each α_{i0} and $\boldsymbol{\beta}_0$. Condition (A5) puts some restrictions on the conditional density and is to some extent standard in the QR literature. Compare Assumption 4 in Horowitz (1998) and condition (ii) of Theorem 3 in Angrist, Chernozhukov, and Fernandez-Val (2006). We require no less than three times differentiability of the conditional density in order to make the smoothing bias negligible.

Condition (A6) corresponds to Assumption 5 in Horowitz (1998). Condition (A6) (b) requires $K(\cdot)$ to be a higher order kernel. Conditions (A7)-(A9) are only used in the proof of Theorem 3.1. The requirement that $a < 1/3$ is explained in Horowitz (1998). The proof of Theorem 3.1 below shows that the asymptotic bias of $\hat{\boldsymbol{\beta}}$ is $\max\{O(T^{-1}), O(h_n^r)\}$, where the $O(T^{-1})$ term corresponds to the estimation error of each individual effect and the $O(h_n^r)$ term corresponds to the smoothing bias. To make $h_n^r = o(T^{-1})$, we need $ar > 1$. A higher order kernel is needed to ensure condition (A7). Several higher order kernels are listed in Muller (1984). An implicit assumption behind this condition is that T grows as fast as n. In practice, it is recommended to set $h_n = o\{(nT)^{-1/(2r)}\}$ in order to make the smoothing bias $o\{(nT)^{-1/2}\}$. Conditions (A8) and (A9) are concerned with the asymptotic covariance matrix and the asymptotic bias, respectively.

3.2 Main results

We first show consistency of the estimator. Although the main concern is the distributional result, consistency of the estimator is an important prerequisite since it guarantees that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is an interior point of $\mathcal{A}^n \times \mathcal{B}$ and thus satisfies the first order condition (2.4) with probability approaching one. As in Kato, Galvao, and Montes-Rojas (2010), we say that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is weakly consistent if $\max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}| \xrightarrow{p} 0$ and $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. The proof of Proposition 3.1 is relegated to Appendix B.1.

Proposition 3.1. *Assume that* $\log n/T \to 0$ *and* $h_n \to 0$. *Then, under conditions (A1)-(A6),* $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *is weakly consistent.*

The condition on T for the consistency in general depends on the order of $\boldsymbol{x}_{it}$'s

[4] Hahn and Kuersteiner (2002, 2004) studied the bias correction for dynamic panel data models with smooth objective functions. The extension of their results to the current case is conceptually straightforward but technically involved.

moment. See Theorem 2.1 of Kato, Galvao, and Montes-Rojas (2010).

Next, we present the limiting distribution of the FE-SQR estimator when n and T grow at the same rate. The proof of Theorem 3.1 is relegated to Appendix B.2.

Theorem 3.1. *Assume that $n/T \to \rho$ for some $\rho > 0$. Then, under conditions (A1)-(A9), we have*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\{\sqrt{\rho}\boldsymbol{b}, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}\}, \tag{3.1}$$

where

$$\boldsymbol{b} := \Gamma^{-1}\left[\frac{\tau(1-\tau)}{2}\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} s_i^2 \boldsymbol{\nu}_i\right)\right]. \tag{3.2}$$

The bias depends on the conditional and unconditional densities and their first derivatives. This may be naturally expected because of the nature of QR. On the other hand, it is perhaps surprising that the asymptotic bias is of a simple form in comparison with the general bias formula displayed in Hahn and Newey (2004). The proof shows that the bias has actually three sources.[5] The first comes from the correlation between $\hat{\alpha}_i$ and the α_i-derivative of the α_i-score evaluated at the truth. The second comes from the correlation between $\hat{\alpha}_i$ and the α_i-derivative of the $\boldsymbol{\beta}$-score evaluated at the truth. The third comes from the variance of $\hat{\alpha}_i$. It is shown that the first and second are canceled out at the $O(T^{-1})$ rate and only the third one remains in the final bias expression.

Our proof strategy is substantially different from a functional expansion method of Li, Lindsay, and Waterman (2003) because it would be unsuitable to handle the situation where the kernel smoothing is involved. The proof strategy is perhaps closer in spirit to the iterative method of stochastic expansion displayed in Rilstone, Srivastava, and Ullah (1996), but still quite different from theirs. It should be pointed out that although the objective function is smooth in the present problem, the proof is still non-trivial since we have to control the smoothing effect and at the same time handle the problem of diverging number of nuisance parameters.

As stated in the literature, the problem of the limiting distribution of $\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ not being centered is that confidence intervals based on the asymptotic approximation will be incorrect. In particular, even if $\boldsymbol{b}$ is small, the asymptotic bias can be of moderate size when the ratio n/T is large. In what follows, we shall consider the bias correction to the FE-SQR estimator.

We first consider a simple one-step bias correction based on the analytic form of the asymptotic bias. Put $\hat{u}_{it} := y_{it} - \hat{\alpha}_i - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}$. Then, $s_i, \boldsymbol{\gamma}_i, \boldsymbol{\nu}_i$ and Γ can be

[5]These do not match one by one what Hahn and Newey (2004) called the sources of the bias because of the difference of the proof strategies.

nonparametrically estimated by

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

$$\hat{\boldsymbol{\nu}}_i := \frac{1}{Th_n^2}\sum_{t=1}^{T} K^{(1)}(\hat{u}_{it}/h_n)(\boldsymbol{x}_{it} - \hat{\boldsymbol{\gamma}}_i), \ \hat{\Gamma}_n := \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{x}_{it}(\boldsymbol{x}_{it}' - \hat{\boldsymbol{\gamma}}_i'),$$

where $K^{(1)}(u) = dK(u)/du$. With these estimates, we can estimate $\boldsymbol{b}$ by

$$\hat{\boldsymbol{b}} := \hat{\Gamma}_n^{-1}\left[\frac{\tau(1-\tau)}{2} \cdot \frac{1}{n}\sum_{i=1}^{n} \hat{s}_i^2 \hat{\boldsymbol{\nu}}_i\right].$$

Then, the one-step bias corrected estimator is defined as $\hat{\boldsymbol{\beta}}^1 := \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{b}}/T$. In practice, there is no need to use the same kernel and the same bandwidth to estimate $\boldsymbol{\beta}_0$ and $\boldsymbol{b}$. In particular, the kernel may not be higher order to estimate $\boldsymbol{b}$.

The next theorem shows that under the same conditions as Theorem 3.1, $\hat{\boldsymbol{\beta}}^1$ has the limiting normal distribution with mean zero and the same covariance matrix as $\hat{\boldsymbol{\beta}}$. The proof of the theorem is relegated to Appendix B.3.

Theorem 3.2. *Under the same conditions as Theorem 3.1, we have*

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

Remark 3.1 (Estimation of the asymptotic covariance matrix)**.** As a byproduct of Theorem 3.2, we can obtain a consistent estimator of the asymptotic covariance matrix. The proof of Theorem 3.2 shows that $\hat{\Gamma}_n \xrightarrow{p} \Gamma$ and $\hat{\boldsymbol{\gamma}}_i \xrightarrow{p} \boldsymbol{\gamma}_i$ uniformly over $1 \leq i \leq n$. Thus, the matrix V can be consistently estimated by

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

so under the same conditions as Theorem 3.1, it is shown that $\tau(1-\tau)\hat{\Gamma}_n^{-1}\hat{V}_n\hat{\Gamma}_n^{-1}$ is consistent for the asymptotic covariance matrix $\tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}$.

Under suitable regularity conditions, the mean-zero asymptotic normality of the bias corrected MLE for smooth likelihood functions hold when $n/T^3 \to 0$ (see Theorem 2 of Hahn and Newey, 2004). This is because the leading term of the bias expansion is $O_p(T^{-1})$ and the next term is $O_p(T^{-2})$. In our case, it is important to note that the conclusion of Theorem 3.2 may hold when T grows with a slower rate than n. However, the second term of the bias expansion for $\hat{\boldsymbol{\beta}}$ is not $O_p(T^{-2})$ and a precise condition on T to ensure the conclusion of Theorem 3.2 depends on the order of the kernel and the decreasing rate of the bandwidth, and is not of a "clean" form because of the presence of kernel smoothing. Thus, we state the theorem in the present form.

We next consider the half-panel jackknife method originally proposed by Dhaene and Jochmans (2009), which is an automatic way of removing the bias of $\hat{\boldsymbol{\beta}}$. Suppose for a moment that T is even. Partition $\{1,\dots,T\}$ into two subsets, $S_1 := \{1,\dots,T/2\}$ and $S_2 := \{T/2+1,\dots,T\}$. Let $\hat{\boldsymbol{\beta}}_{S_a}$ be the FE-SQR estimate based on the data $\{(y_{it},\boldsymbol{x}_{it}), 1 \le i \le n, t \in S_a\}$ for $a = 1,2$. The half-panel jackknife estimator is defined as $\hat{\boldsymbol{\beta}}_{1/2} := 2\hat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}_{1/2}$, where $\bar{\boldsymbol{\beta}}_{1/2} := (\hat{\boldsymbol{\beta}}_{S_1} + \hat{\boldsymbol{\beta}}_{S_2})/2$. For simplicity, suppose for a moment that we use the same bandwidth to construct $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{S_a}$ $(a = 1,2)$.

Theorem 3.3. *Under the same conditions as Theorem 3.1, we have*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{1/2} - \boldsymbol{\beta}_0) \xrightarrow{d} N\{\boldsymbol{0}, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}\}.$$

The proof of this theorem is immediate from the asymptotic representation of the FE-SQR estimator that appears in the proof of Theorem 3.1. The half-panel jackknife estimator would be attractive in both theoretical and practical senses. In fact, its validity does not require any extra condition as we can see from Theorem 3.3. It would be also preferable from a practical point of view since it does not require the nonparametric estimation of the bias term and at the same time is easy to implement.

One can find some other options to correct the bias of the FE-SQR estimator from the literature (see, for instance, Hahn and Newey, 2004; Arellano and Hahn, 2005; Bester and Hansen, 2009; Dhaene and Jochmans, 2009). Though it is an interesting topic, the exhaustive comparative study of such methods is beyond the scope of this paper, and we would restrict our attention to the one-step bias correction and the half-panel jackknife method.

We end this section with an example that will give an intuition for the bias term. This example shows that in some cases like linear location models, the bias term is exactly zero, while in heteroscedastic models, the bias term is in general not zero.

Example 3.1. Consider a linear location-scale model

$$y_{it} = \eta_{i0} + \boldsymbol{x}_{it}'\boldsymbol{\theta}_0^{(1)} + \sigma_{it}\epsilon_{it},$$

where each η_{i0} is a fixed constant, $\sigma_{it} := 1 + \boldsymbol{x}_{it}'\boldsymbol{\theta}_0^{(2)} > 0$, $\{(\epsilon_{it},\boldsymbol{x}_{it}),\ t \ge 1\}$ is i.i.d. for each fixed i and independent across i, and ϵ_{it} has the identical distribution F_ϵ. In this case, $\alpha_{i0} = \eta_{i0} + F_\epsilon^{-1}(\tau)$, $\boldsymbol{\beta}_0 = \boldsymbol{\theta}_0^{(1)} + F_\epsilon^{-1}(\tau)\boldsymbol{\theta}_0^{(2)}$ and $f_i(u|\boldsymbol{x}_{i1}) = \sigma_{it}^{-1} f_\epsilon(F_\epsilon^{-1}(\tau) + \sigma_{it}^{-1}u)$, where f_ϵ is the density of F_ϵ. Then, a simple algebra yields that

$$s_i = \frac{1}{\mathrm{E}[\sigma_{i1}^{-1}]f_\epsilon(F_\epsilon^{-1}(\tau))},\ \boldsymbol{\nu}_i = f_\epsilon^{(1)}(F_\epsilon^{-1}(\tau))\left\{\frac{\mathrm{E}[\sigma_{i1}^{-2}]\mathrm{E}[\sigma_{i1}^{-1}\boldsymbol{x}_{i1}]}{\mathrm{E}[\sigma_{i1}^{-1}]} - \mathrm{E}[\sigma_{i1}^{-2}\boldsymbol{x}_{i1}]\right\}.$$

From these expressions, we can see that the bias vanishes in some cases. For example, if $\sigma_{it} = 1$, i.e., the scale does not depend on $\boldsymbol{x}_{it}$, $\boldsymbol{\nu}_i = \boldsymbol{0}$ for all i and thus $\boldsymbol{b} = \boldsymbol{0}$. Even if the scale depends on $\boldsymbol{x}_{it}$, the bias is zero when $f_\epsilon^{(1)}(F_\epsilon^{-1}(\tau)) = 0$, which occurs, for example, when the median and the mode of F_ϵ are zero and $\tau = 0.5$.

4 A Monte Carlo study

4.1 Design

In this section, we report a small Monte Carlo study to investigate the finite sample performance of the bias correction. A simple version of model (2.1) is considered in this study:

$$y_{it} = \eta_i + x_{it} + (1 + 0.2x_{it})\epsilon_{it}, \tag{4.1}$$

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 $F = N(0,1)$ or χ^2_3.[6] In this model, $\alpha_i^* = \alpha_i^*(\tau) = \eta_i + F^{-1}(\tau)$ and $\beta_0 = \beta_0(\tau) = 1 + 0.2F^{-1}(\tau)$. We let $n \in \{100, 200\}$, $T \in \{8, 12, 16, 20\}$ and $\tau \in \{0.25, 0.50, 0.75\}$. As in Horowitz (1998), we used the fourth order kernel

$$K(u) := \frac{105}{64}(1 - 5u^2 + 7u^4 - 3u^6)I(|u| \le 1).$$

The number of repetitions for each Monte Carlo experiment is 2,000. The FE-SQR objective function is not necessarily convex and may have several local optima. Thus, the choice of the starting value is important when computing the estimate. We chose the standard QR estimate as a reasonable starting value. We used `quantreg` package (Koenker, 2009) to compute the standard QR estimates. The estimators under consideration are the standard QR estimator $\hat{\beta}_{\mathrm{KB}}$ defined by (2.2); the FE-SQR estimator $\hat{\beta}$; the one-step bias corrected FE-SQR estimator $\hat{\beta}^1$ and the half-panel jackknife estimator $\hat{\beta}_{1/2}$ described in Section 3.

The choice of the bandwidth is a common, sometimes difficult problem when the kernel smoothing is used. In this experiment, we used $h_{1n} = c_1 s_1 (nT)^{-1/7}$ to construct the FE-SQR estimate, where c_1 is some constant and s_1 is the sample standard deviation of $\tilde{u}_{it} = y_{it} - \hat{\alpha}_{i,\mathrm{KB}} - x_{it}\hat{\beta}_{\mathrm{KB}}$, and $h_{2n} = c_2 s_2 T^{-1/5}$ to construct the bias estimate, where c_2 is some constant and s_2 is the sample standard deviation of $\hat{u}_{it} = y_{it} - \hat{\alpha}_i - x_{it}\hat{\beta}$.[7] These choices are only for simplicity. We leave the choice of the bandwidth as a future topic. Some intuition behind these choices, however, may be explained as follows. According to condition (A7) and the discussion in Section 3.1, the bandwidth used to construct the FE-SQR estimate should be of the same order as $(nT)^{-a/2}$ with $1/3 < a < 1/4$ when n and T grow at the same rate. Our choice h_{1n} satisfies this restriction except for the fact that it depends on the data.[8] On the other

[6] In fact, we examined the cases with a different degree of heteroscedasticity and a different error distribution, but the results are qualitatively similar.

[7] $(\hat{\alpha}_{1,\mathrm{KB}}, \ldots, \hat{\alpha}_{n,\mathrm{KB}}, \hat{\beta}_{\mathrm{KB}})$ is defined as a solution to (2.2) in the present context. Recall that we used the standard QR estimate as a starting value to compute the FE-SQR estimate.

[8] Though it is an important topic, we do not further discuss the validity of the data dependent bandwidth.

hand, in view of the proof of Theorem 3.2, the bandwidth used to construct $\hat{\boldsymbol{b}}$, say h_{2n}, should be such that $h_{2n} \asymp T^{-a}$ with $0 < a < 1/3$ to make $\hat{\boldsymbol{b}}$ consistent for $\boldsymbol{b}$. Our choice h_{2n} satisfies this restriction. The reason for the use of the different bandwidths is that in practical situations where n is much larger than T, the choice h_{1n} is inclined to be rather small for $\hat{\boldsymbol{b}}$, which results in less stability of the analytic bias correction. Regarding the constants in h_{1n} and h_{2n}, we set $c_1 = 1$ and $c_2 = 2$ for specificity.

When T is small, on rare occasions $\hat{f}_i$ is very close to zero for some i, which results in an unreasonable value of $\hat{\beta}^1$. To avoid this, we devised the trimming. Put $\hat{I}_i := I(\hat{f}_i > c)$ for some small constant $c > 0$. We modified $\hat{\Gamma}_n$ and $\hat{b}$ as

$$\hat{\Gamma}_n = \frac{1}{nT}\sum_{i=1}^{n}\hat{I}_i\sum_{t=1}^{T}K_{h_{2n}}(\hat{u}_{it})x_{it}(x_{it}-\hat{\gamma}_i),\ \hat{b} = \hat{\Gamma}_n^{-1}\left[\frac{\tau(1-\tau)}{2}\cdot\frac{1}{n}\sum_{i=1}^{n}\hat{I}_i\hat{s}_i^2\hat{\nu}_i\right].$$

In our implementation, we set $c = 0.01$ for simplicity. The trimming is frequently used in the nonparametric and semiparametric literature. We refer to Section 6.4 of Ichimura and Todd (2007) for discussion on trimming.

4.2 Results

Before looking at the Monte Carlo results, it is worthwhile to summarize our objects of this experiment. The objects are twofold. The first one is to examine whether the bias correction methods work for finite sample. In view of the theoretical results, the bias of $\hat{\beta}^1$ and $\hat{\beta}_{1/2}$ should be smaller than $\hat{\beta}$, at least for moderate T. In fact, the methods described in Section 3.2 are intended to remove the bias of order $O(T^{-1})$, so T times the bias of $\hat{\beta}^1$ and $\hat{\beta}_{1/2}$ should decrease as T increases. To this end, as in Phillips and Sul (2007), we computed T times the bias of each estimator. The second object is to examine the effect of the bias correction to the variance of the estimator. The theoretical results suggest that the asymptotic variance of $\hat{\beta}^1$ or $\hat{\beta}_{1/2}$ is the same as that of $\hat{\beta}$. To this end, we computed the standard deviation of each estimator. The results for $\hat{\beta}_{\text{KB}}$ are only for reference, and we mainly focus on the comparison between $\hat{\beta}$, $\hat{\beta}^1$ and $\hat{\beta}_{1/2}$.[9]

Table 1 presents the results for the normal error with $n = 100$. It can be observed that the one-step bias correction works well for each τ in terms of the bias reduction. In particular, T times the bias of $\hat{\beta}^1$ roughly decreases as T increases, while that of $\hat{\beta}$ (and $\hat{\beta}_{\text{KB}}$) behaves like stationary expect for the $\tau = 0.5$ case. For $\tau = 0.5$, as we have seen in Example 3.1, the order $O(T^{-1})$ bias of the FE-SQR estimator is exactly zero when ϵ_{it} has the standard normal distribution, so it is not surprising that the bias of $\hat{\beta}^1$ is sightly larger than that of $\hat{\beta}$, and T times the bias of $\hat{\beta}$ also decreases as T

[9]Recall that it is not known whether a result analogous to Theorem 3.1 holds for $\hat{\beta}_{\text{KB}}$.

increases in that case. It is curious to note that the half panel jackknife method does not work for this and other cases reported here. On the other hand, the one-step bias correction slightly increases the standard deviation. This may reflect the fact that the bias estimate is likely to be unstable in the finite sample (especially when T is small) since it depends on the nonparametric estimation of the conditional and unconditional densities and their first derivatives.

Table 2 presents the results for the χ_2^3 error with $n = 100$. In this case, the one-step bias correction does not work so well as in the previous case in particular for $\tau = 0.25$. For $\tau = 0.25$, the bias of $\hat{\beta}$ is very small, so the bias correction might be redundant in that case. For $\tau = 0.5$, the bias of $\hat{\beta}^1$ is larger than that of $\hat{\beta}$ for $T = 8, 12$ but is smaller than that of $\hat{\beta}$ for $T = 16, 20$. The bias of $\hat{\beta}$ is the largest when $\tau = 0.75$ among the quantiles under consideration. It is noticeable that the one-step bias correction works quite well in that case. The results for the standard deviation are analogous to the previous case, Table 1.

The results for $n = 200$, presented in Tables 3 (normal error) and 4 (χ_3^2 error), basically parallel those for $n = 100$. Overall, in our limited examples, we may conclude that the one-step bias correction is able to substantially reduce the bias in many cases, although it slightly increases the variability in the small sample partly because of the nonparametric estimation of the bias term.

5 Concluding Remarks

In this paper, we have shown that the smoothed quantile regression estimator for panel QR models has a limiting normal distribution with a bias in the mean when n and T grow at the same rate. We have further considered the two methods of correcting the bias of the estimator, namely an analytic bias correction and the half-panel jackknife method of Dhaene and Jochmans (2009), and theoretically shown that they actually eliminate the bias in the limiting distribution. The contribution of this paper is to open a new door to the bias correction of common parameters' estimators in panel QR models, which we believe is an important research area.

A substantial difference from smooth nonlinear panel models is that the FE-SQR estimator requires the choice of the bandwidth and its asymptotic bias consists of non-parametric objects. In particular, the estimation of the bias term is more complicated than in smooth nonlinear panel models. As stated in Section 4, the choice of the band-width is an important topic to be further investigated in a future work. Moreover, a thorough finite sample comparison of various bias correction methods and inference procedures for panel QR models would be an additional interesting future research topic.[10]

A Fixed-T inconsistency of the standard QR estimator

In this section, we present an example in which the standard QR estimator of the common parameter is actually inconsistent as n goes to infinity while T is fixed. Suppose as before that $\tau \in (0,1)$ is fixed. Let us consider a simple model

$$y_{it} = 1 + d_{it} + (1 + d_{it})\epsilon_{it},\ i = 1,\dots,n; T = 1,2.$$

Assume that (ϵ_{it}, d_{it}) is i.i.d. across i and t; $\{\epsilon_{it}\}$ and $\{d_{it}\}$ are independent; $\mathrm{P}(d_{it} = 0) = \mathrm{P}(d_{it} = 1) = 1/2$ and $\epsilon_{it} \sim N(0,1)$. In this case, the pair (y_{it}, d_{it}) satisfies the conditional quantile restriction

$$Q_\tau(y_{it}|d_{it}) = \{1 + \Phi^{-1}(\tau)\} + \{1 + \Phi^{-1}(\tau)\}d_{it} =: \alpha_{i0} + \beta_0 d_{it},$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. We consider the consequence of applying the standard QR estimation to this model.

According to Graham, Hahn, and Powell (2009), $\hat{\beta}_{\mathrm{KB}}$ satisfies

$$\sum_{i=1}^n |(y_{i2} - y_{i1}) - \hat{\beta}_{\mathrm{KB}}(d_{i2} - d_{i1})| = \min_\beta \sum_{i=1}^n |(y_{i2} - y_{i1}) - \beta(d_{i2} - d_{i1})|,$$

that is, $\hat{\beta}_{\mathrm{KB}}$ is the least absolute deviation estimator with $y_{i2} - y_{i1}$ as a response variable and $d_{i2} - d_{i1}$ as an explanatory variable. From Angrist, Chernozhukov, and Fernandez-Val (2006), $\hat{\beta}_{\mathrm{KB}}$ converges in probability to the unique β^* that solves

$$\mathrm{E}[\{I(y_{i2} - y_{i1} \le \beta^*(d_{i2} - d_{i1})) - 1/2\}(d_{i2} - d_{i1})] = 0.$$

We shall compute β^*. If $d_{i2} = 1$ and $d_{i1} = 0$, then $y_{i2} - y_{i1} = 1 + 2\epsilon_{i2} - \epsilon_{i1} \sim N(1,5)$. Thus, $\mathrm{E}[\{I(y_{i2} - y_{i1} \le \beta^*(d_{i2} - d_{i1})) \mid d_{i2} = 1, d_{i1} = 0] = \Phi((\beta^* - 1)/\sqrt{5})$. On the other hand, if $d_{i2} = 0$ and $d_{i1} = 1$, then $y_{i2} - y_{i1} = -1 + \epsilon_{i2} - 2\epsilon_{i1} \sim N(-1,5)$. Thus, $\mathrm{E}[\{I(y_{i2} - y_{i1} \le \beta^*(d_{i2} - d_{i1})) \mid d_{i2} = 0, d_{i1} = 1] = 1 - \Phi((\beta^* - 1)/\sqrt{5})$. Therefore, we have

$$\mathrm{E}[\{I(y_{i2} - y_{i1} \le \beta^*(d_{i2} - d_{i1})) - 1/2\}(d_{i2} - d_{i1})] = \frac{1}{2}\Phi\left(\frac{\beta^* - 1}{\sqrt{5}}\right) - \frac{1}{4},$$

which takes zero only when $\beta^* = 1$, so $\hat{\beta}_{\mathrm{KB}}$ is inconsistent unless $\tau = 1/2$. The absolute bias of $\hat{\beta}_{\mathrm{KB}}$ is $|\Phi^{-1}(\tau)|$, which diverges as τ goes to zero or one.

[10] Buchinsky (1995) compared several inference methods for QR models with cross section data through extensive simulations.

B Proofs

For any triangular sequence $\{a_{ni}\}_{i=1}^n$ and $\epsilon_n \to 0$, we write $a_{ni} = \bar{o}(\epsilon_n)$ if $\max_{1\le i\le n} |a_{ni}| = o(\epsilon_n)$. We also define $\bar{O}(\cdot)$, $\bar{o}_p(\cdot)$ and $\bar{O}_p(\cdot)$ in a similar fashion.

B.1 Proof of Proposition 3.1

Put $\mathbb{M}_{ni}(\alpha, \boldsymbol{\beta}) := T^{-1}\sum_{t=1}^T (y_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})\{\tau - G_{h_n}(y_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})\}$, $\tilde{\mathbb{M}}_{ni}(\alpha, \boldsymbol{\beta}) := T^{-1}\sum_{t=1}^T (y_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})\{\tau - I(y_{it} \le \alpha + \boldsymbol{x}_{it}'\boldsymbol{\beta})\}$ and $\Delta_i(\alpha, \boldsymbol{\beta}) := \mathrm{E}[\tilde{\mathbb{M}}_{ni}(\alpha, \boldsymbol{\beta}) - \tilde{\mathbb{M}}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)]$. Without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. Let $\|\cdot\|_1$ denote the ℓ_1 norm. For $\delta > 0$, define $N_\delta := \{(\alpha, \boldsymbol{\beta}) : |\alpha| + \|\boldsymbol{\beta}\|_1 \le \delta\}$ and N_δ^c by its complement in $\mathbb{R}^{p+1}$. We first show that for any $\delta > 0$,

$$\inf_{i\ge 1} \min_{(\alpha,\boldsymbol{\beta})\in N_\delta^c} \Delta_i(\alpha, \boldsymbol{\beta}) > 0. \tag{B.1}$$

To see this, use the identity of Knight (1998) to obtain

$$\Delta_i(\alpha, \boldsymbol{\beta}) = \mathrm{E}\left[\int_0^{\alpha + \boldsymbol{x}_{i1}'\boldsymbol{\beta}} \{F_i(s|\boldsymbol{x}_{i1}) - \tau\} ds\right],$$

from which we can see that each map $(\alpha, \boldsymbol{\beta}) \mapsto \Delta_i(\alpha, \boldsymbol{\beta})$ is convex by differentiation. By condition (A5), we can expand $\Delta_i(\alpha, \boldsymbol{\beta})$ uniformly over $(\alpha, \boldsymbol{\beta})$ such that $|\alpha| + \|\boldsymbol{\beta}\|_1 = \delta$ and $i \ge 1$ as

$$\Delta_i(\alpha, \boldsymbol{\beta}) = (\alpha, \boldsymbol{\beta}')\mathrm{E}[f_i(0|\boldsymbol{x}_{i1})(1, \boldsymbol{x}_{i1}')'(1, \boldsymbol{x}_{i1}')](\alpha, \boldsymbol{\beta}')' + o(\delta^2),\ \delta \to 0.$$

By condition (A4), there exist a positive constant c independent of i such that

$$(\alpha, \boldsymbol{\beta}')\mathrm{E}[f_i(0|\boldsymbol{x}_{i1})(1, \boldsymbol{x}_{i1}')'(1, \boldsymbol{x}_{i1}')](\alpha, \boldsymbol{\beta}')' \ge c(|\alpha| + \|\boldsymbol{\beta}\|_1)^2,\ \forall i \ge 1.$$

Thus, there exists a positive constant δ_0 such that for any $0 < \delta \le \delta_0$,

$$\inf_{i\ge 1} \min_{|\alpha|+\|\boldsymbol{\beta}\|_1=\delta} \Delta_i(\alpha, \boldsymbol{\beta}) > 0. \tag{B.2}$$

Now, pick any $0 < \delta \le \delta_0$ and any $(\alpha, \boldsymbol{\beta}) \in N_\delta^c$. Take $r = \delta/(|\alpha| + \|\boldsymbol{\beta}\|_1)$, $\tilde{\alpha} = r\alpha$ and $\tilde{\boldsymbol{\beta}} = r\boldsymbol{\beta}$, so that $|\tilde{\alpha}| + \|\tilde{\boldsymbol{\beta}}\|_1 = \delta$. Because of the convexity of the map $(\bar{\alpha}, \bar{\boldsymbol{\beta}}) \mapsto \Delta_i(\bar{\alpha}, \bar{\boldsymbol{\beta}})$, we have $\Delta_i(\alpha, \boldsymbol{\beta}) \ge r^{-1}\Delta_i(\tilde{\alpha}, \tilde{\boldsymbol{\beta}})$. Thus, by (B.2), (B.1) holds for $0 < \delta \le \delta_0$. It is then obvious that (B.1) holds for any $\delta > 0$ since the left hand side of (B.1) is nondecreasing in $\delta > 0$.

Next, we shall show that

$$\max_{1\le i\le n} \sup_{(\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}} |\mathbb{M}_{ni}(\alpha, \boldsymbol{\beta}) - \mathbb{M}_{ni}(0, \mathbf{0}) - \Delta_i(\alpha, \boldsymbol{\beta})| \xrightarrow{p} 0. \tag{B.3}$$

Observe that

$$
\begin{aligned}
&\mathbb{M}_{ni}(\alpha,\boldsymbol{\beta}) - \mathbb{M}_{ni}(0,\mathbf{0}) - \Delta_i(\alpha,\boldsymbol{\beta}) \\
&= \{\mathbb{M}_{ni}(\alpha,\boldsymbol{\beta}) - \tilde{\mathbb{M}}_{ni}(\alpha,\boldsymbol{\beta})\} - \{\mathbb{M}_{ni}(0,\mathbf{0}) - \tilde{\mathbb{M}}_{ni}(0,\mathbf{0})\} \\
&\quad + \{\tilde{\mathbb{M}}_{ni}(\alpha,\boldsymbol{\beta}) - \tilde{\mathbb{M}}_{ni}(0,\mathbf{0}) - \Delta_i(\alpha,\boldsymbol{\beta})\}.
\end{aligned}
$$

From the proof of Lemma 1 in Horowitz (1998), the first and second terms are $O(h_n)$ uniformly over $(\alpha,\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{B}$ and $1 \le i \le n$. From the proof of Theorem 2.1 in Kato, Galvao, and Montes-Rojas (2010), when $\log n/T \to 0$, the third term is $o_p(1)$ uniformly over $(\alpha,\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{B}$ and $1 \le i \le n$. Thus, we obtain (B.3).

We shall show $\hat{\boldsymbol{\beta}} \xrightarrow{p} \mathbf{0} = \boldsymbol{\beta}_0$. Pick any $\delta > 0$ and put $\epsilon := \inf_{i\ge 1}\min_{(\alpha,\boldsymbol{\beta})\in N_\delta^c}\Delta_i(\alpha,\boldsymbol{\beta}) > 0$. Observe that when $\|\hat{\boldsymbol{\beta}}\|_1 > \delta$,

$$
\begin{aligned}
&n^{-1}\textstyle\sum_{i=1}^n\{\mathbb{M}_{ni}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(0,\mathbf{0})\} \\
&\ge n^{-1}\textstyle\sum_{i=1}^n \min_{(\alpha,\boldsymbol{\beta})\in N_\delta^c\cap\mathcal{A}\times\mathcal{B}}\{\mathbb{M}_{ni}(\alpha,\boldsymbol{\beta}) - \mathbb{M}_{ni}(0,\mathbf{0})\} \\
&\ge n^{-1}\textstyle\sum_{i=1}^n \min_{(\alpha,\boldsymbol{\beta})\in N_\delta^c\cap\mathcal{A}\times\mathcal{B}}\{\mathbb{M}_{ni}(\alpha,\boldsymbol{\beta}) - \mathbb{M}_{ni}(0,\mathbf{0}) - \Delta_i(\alpha,\boldsymbol{\beta})\} \\
&\quad + \inf_{i\ge 1}\min_{(\alpha,\boldsymbol{\beta})\in N_\delta^c}\Delta_i(\alpha,\boldsymbol{\beta}) \\
&\ge -\max_{1\le i\le n}\sup_{(\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}}|\mathbb{M}_{ni}(\alpha,\boldsymbol{\beta}) - \mathbb{M}_{ni}(0,\mathbf{0}) - \Delta_i(\alpha,\boldsymbol{\beta})| + \epsilon.
\end{aligned}
$$

By definition, $n^{-1}\sum_{i=1}^n \mathbb{M}_{ni}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) \le n^{-1}\sum_{i=1}^n \mathbb{M}_{ni}(0,\mathbf{0})$. Thus,

$$
\mathrm{P}(\|\hat{\boldsymbol{\beta}}\|_1 > \delta) \le \mathrm{P}\left\{\max_{1\le i\le n}\sup_{(\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}}|\mathbb{M}_{ni}(\alpha,\boldsymbol{\beta}) - \mathbb{M}_{ni}(0,\mathbf{0}) - \Delta_i(\alpha,\boldsymbol{\beta})| \ge \epsilon\right\} \to 0,
$$

which implies that $\hat{\boldsymbol{\beta}} \xrightarrow{p} \mathbf{0}$.

We next prove $\max_{i\le i\le n}|\hat{\alpha}_i| \xrightarrow{p} 0$. Pick any $\delta > 0$ and put ϵ as before. Recall that $\hat{\alpha}_i = \arg\min_{\alpha\in\mathcal{A}}\mathbb{M}_{ni}(\alpha,\hat{\boldsymbol{\beta}})$. When $|\hat{\alpha}_i| > \delta$ for some $1 \le i \le n$,

$$
\begin{aligned}
&\mathbb{M}_{ni}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(0,\hat{\boldsymbol{\beta}}) \\
&= \min_{|\alpha|>\delta,\alpha\in\mathcal{A}}\{\mathbb{M}_{ni}(\alpha,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(0,\hat{\boldsymbol{\beta}})\} \\
&= \min_{|\alpha|>\delta,\alpha\in\mathcal{A}}\{\mathbb{M}_{ni}(\alpha,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(0,\mathbf{0})\} - \{\mathbb{M}_{ni}(0,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(0,\mathbf{0})\} \\
&\ge \min_{|\alpha|>\delta,\alpha\in\mathcal{A}}\{\mathbb{M}_{ni}(\alpha,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(0,\mathbf{0}) - \Delta_i(\alpha,\hat{\boldsymbol{\beta}})\} + \epsilon \\
&\quad - \{\mathbb{M}_{ni}(0,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(0,\mathbf{0}) - \Delta_i(0,\hat{\boldsymbol{\beta}})\} - \Delta_i(0,\hat{\boldsymbol{\beta}}) \\
&\ge -2\max_{1\le j\le n}\sup_{(\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}}|\mathbb{M}_{nj}(\alpha,\boldsymbol{\beta}) - \mathbb{M}_{nj}(0,\mathbf{0}) - \Delta_j(\alpha,\boldsymbol{\beta})| + \epsilon - \max_{1\le j\le n}\Delta_j(0,\hat{\boldsymbol{\beta}}).
\end{aligned}
$$

Thus, we obtain the inclusion relation

$$\begin{aligned}
&\{|\hat{\alpha}_i| > \delta,\ 1 \le \exists i \le n\} \\
&\subset \left\{ \max_{1\le i\le n} \sup_{(\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}} |\mathbb{M}_{ni}(\alpha,\boldsymbol{\beta}) - \mathbb{M}_{ni}(0,\mathbf{0}) - \Delta_i(\alpha,\boldsymbol{\beta})| \ge \frac{\epsilon}{3} \right\} \cup \left\{ \max_{1\le i\le n} \Delta_i(0,\hat{\boldsymbol{\beta}}) \ge \frac{\epsilon}{3} \right\} \\
&=: A_{1n} \cup A_{2n}.
\end{aligned}$$

By (B.3), $\mathrm{P}(A_{1n}) \to 0$. On the other hand, since $\Delta_i(0,\hat{\boldsymbol{\beta}}) \le 2M\|\hat{\boldsymbol{\beta}}\|$, consistency of $\hat{\boldsymbol{\beta}}$ implies that $\mathrm{P}(A_{2n}) \to 0$. Therefore, we complete the proof. □

B.2 Proof of Theorem 3.1

Recall the definition of $\mathbb{H}_{ni}^{(1)}(\alpha_i,\boldsymbol{\beta})$ and $\mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha},\boldsymbol{\beta})$ in Section 2. Put $H_{ni}^{(1)}(\alpha_i,\boldsymbol{\beta}) := \mathrm{E}[\mathbb{H}_{ni}^{(1)}(\alpha_i,\boldsymbol{\beta})]$ and $H_{n}^{(2)}(\boldsymbol{\alpha},\boldsymbol{\beta}) := \mathrm{E}[\mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha},\boldsymbol{\beta})]$. The proof of Theorem 3.1 consists of a series of lemmas. Throughout the proof, we assume all the conditions of Theorem 3.1. Lemmas B.1-B.3 below are used to derive a convenient expansion of $\hat{\boldsymbol{\beta}}$ which will be given in (B.15).

Lemma B.1. *Take $\delta_{1n} \to 0$ and $\delta_{2n} \to 0$ such that* $\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}| = O_p(\delta_{1n})$ *and* $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(\delta_{2n})$. *Put* $d_n := (\delta_{1n} + \delta_{2n})h_n^{r-1} + \delta_{1n}^2 h_n^{r-2} + \delta_{1n}\delta_{2n} + \delta_{2n}^2 + \delta_{1n}^3$. *Then, we have*

$$\begin{aligned}
\hat{\alpha}_i - \alpha_{i0} = {}& s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0) - \boldsymbol{\gamma}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2^{-1} s_i f_i^{(1)}(0)(\hat{\alpha}_i - \alpha_{i0})^2 \\
&+ s_i[\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i},\boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i},\boldsymbol{\beta}_0)\}] \\
&+ \bar{O}_p(d_n), \qquad \text{(B.4)} \\
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = {}& \Gamma_n^{-1}\{-n^{-1}\textstyle\sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\} \\
&- \Gamma_n^{-1} n^{-1} \textstyle\sum_{i=1}^n \boldsymbol{\gamma}_i[\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i},\boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i},\boldsymbol{\beta}_0)\}] \\
&+ \Gamma_n^{-1}[\{\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\} - \{H_n^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - H_n^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\}] \\
&+ \Gamma_n^{-1}\{(2n)^{-1} \textstyle\sum_{i=1}^n \boldsymbol{\nu}_i(\hat{\alpha}_i - \alpha_{i0})^2\} + O_p(d_n). \qquad \text{(B.5)}
\end{aligned}$$

Proof of Lemma B.1. The consistency result shows that $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ satisfies the first order condition (2.4) with probability approaching one. The first order condition implies that

$$\begin{aligned}
0 = {}& \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0) + \{H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i},\boldsymbol{\beta}_0)\} \\
&+ [\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i},\boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i},\boldsymbol{\beta}_0)\}]
\end{aligned}$$

Expanding $H_{ni}^{(1)}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}})$ around $(\alpha_{i0},\boldsymbol{\beta}_0)$ and using Lemma C.1, we obtain (B.4). On the other hand, the first order condition implies that

$$
\begin{aligned}
\mathbf{0} = \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + \{H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} \\
+ [\{\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} - \{H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\}]. \quad \text{(B.6)}
\end{aligned}
$$

Expanding $H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ around $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and using Lemma C.1, we obtain

$$
\begin{aligned}
&H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \\
&= -\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}](\hat{\alpha}_i - \alpha_{i0}) - \left\{\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}']\right\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad - \frac{1}{2n}\sum_{i=1}^{n} \mathrm{E}[f_i^{(1)}(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}](\hat{\alpha}_i - \alpha_{i0})^2 + O_p(d_n). \quad \text{(B.7)}
\end{aligned}
$$

Plugging (B.4) into (B.7) yields that $H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ is expanded as

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n} \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\boldsymbol{\gamma}_i - \Gamma_n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{1}{2n}\sum_{i=1}^{n} \boldsymbol{\nu}_i(\hat{\alpha}_i - \alpha_{i0})^2 \\
&-\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\gamma}_i[\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\}] + O_p(d_n).
\end{aligned}
\quad \text{(B.8)}
$$

Combining (B.8) and (B.6) leads to (B.5). □

Put

$$
\begin{aligned}
I_{ni}^{(1)} &:= \{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\}, \\
I_n^{(2)} &:= \{\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} - \{H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\}, \\
\hat{f}_i &:= T^{-1}\textstyle\sum_{t=1}^{T} K_{h_n}(u_{it}), \ \tilde{f}_i^{(1)} := -(Th_n^2)^{-1}\textstyle\sum_{t=1}^{T} \tilde{K}^{(1)}(u_{it}/h_n), \\
\hat{\boldsymbol{g}}_i &:= T^{-1}\textstyle\sum_{t=1}^{T} K_{h_n}(u_{it})\boldsymbol{x}_{it}, \ \tilde{\boldsymbol{g}}_i^{(1)} := -(Th_n^2)^{-1}\textstyle\sum_{t=1}^{T} \tilde{K}^{(1)}(u_{it}/h_n)\boldsymbol{x}_{it}.
\end{aligned}
$$

Lemma B.2. *We have*

$$
\begin{aligned}
I_{ni}^{(1)} &= -\{(\hat{f}_i - \mathrm{E}[\hat{f}_i]) - h_n(\tilde{f}_i^{(1)} - \mathrm{E}[\tilde{f}_i^{(1)}])\}(\hat{\alpha}_i - \alpha_{i0}) \\
&\quad + \bar{o}_p[\{\log n/(Th_n)^{1/2}\}\delta_{2n} + \{\log n/(Th_n^3)^{1/2}\}\delta_{1n}^2], \quad \text{(B.9)} \\
I_n^{(2)} &= -n^{-1}\textstyle\sum_{i=1}^{n}\{(\hat{\boldsymbol{g}}_i - \mathrm{E}[\hat{\boldsymbol{g}}_i]) - h_n(\tilde{\boldsymbol{g}}_i^{(1)} - \mathrm{E}[\tilde{\boldsymbol{g}}^{(1)}])\}(\hat{\alpha}_i - \alpha_{i0}) \\
&\quad + o_p[\{\log n/(Th_n)^{1/2}\}\delta_{2n} + \{\log n/(Th_n^3)^{1/2}\}\delta_{1n}^2], \quad \text{(B.10)}
\end{aligned}
$$

where δ_{1n} and δ_{2n} are given in Lemma B.1.

Proof of Lemma B.2. We only prove (B.9) since the proof of (B.10) is analogous. Observe that $I_{ni}^{(1)}$ is decomposed as $J_{i1} + J_{i2}$, where

$$
\begin{aligned}
J_{i1} &:= [\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \boldsymbol{\beta}_0)\}], \\
J_{i2} &:= [\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \boldsymbol{\beta}_0) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i, \boldsymbol{\beta}_0) - H_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\}].
\end{aligned}
$$

By Taylor's theorem, $J_{i1} = \{\partial_{\boldsymbol\beta}\mathbb{H}_{ni}^{(1)}(\hat\alpha_i, \tilde{\boldsymbol\beta}_i) - \partial_{\boldsymbol\beta} H_{ni}^{(1)}(\hat\alpha_i, \tilde{\boldsymbol\beta}_i)\}'(\hat{\boldsymbol\beta} - \boldsymbol\beta_0)$, where for each i, $\tilde{\boldsymbol\beta}_i$ is on the line segment between $\hat{\boldsymbol\beta}$ and $\boldsymbol\beta_0$. Now, by Lemma C.2, $\partial_{\boldsymbol\beta}\mathbb{H}_{ni}^{(1)}(\hat\alpha_i, \tilde{\boldsymbol\beta}_i) - \partial_{\boldsymbol\beta} H_{ni}^{(1)}(\hat\alpha, \tilde{\boldsymbol\beta}_i) = \bar o_p\{\log n/(Th_n)^{1/2}\}$, implying that $J_{i1} = \bar o_p[\{\log n/(Th)^{1/2}\}\delta_{2n}]$. Similarly, use Taylor's theorem to obtain

$$
\begin{aligned}
J_{i2} = &-\{(\hat f_i - \mathrm{E}[\hat f_i]) - h_n(\tilde f_i^{(1)} - \mathrm{E}[\tilde f_i^{(1)}])\}(\hat\alpha_i - \alpha_{i0}) \\
&+ \{\partial_{\alpha_i}^2 \mathbb{H}_{ni}^{(1)}(\tilde\alpha_i, \boldsymbol\beta_0) - \partial_{\alpha_i}^2 H_{ni}^{(1)}(\tilde\alpha_i, \boldsymbol\beta_0)\}(\hat\alpha_i - \alpha_{i0})^2, \qquad \text{(B.11)}
\end{aligned}
$$

where for each i, $\tilde\alpha_i$ is on the line segment between $\hat\alpha_i$ and α_{i0}. By Lemma C.2, we have $\partial_{\alpha_i}^2 \mathbb{H}_{ni}^{(1)}(\tilde\alpha_i, \boldsymbol\beta_0) - \partial_{\alpha_i}^2 H_{ni}^{(1)}(\tilde\alpha_i, \boldsymbol\beta_0) = \bar o_p\{\log n/(Th_n^3)^{1/2}\}$, implying that the second term of J_{i2} in (B.11) is $\bar o_p[\{\log n/(Th_n^3)^{1/2}\}\delta_{1n}^2]$. Therefore, we complete the proof. □

Lemma B.3. $\max_{1\le i\le n} |\hat\alpha_i - \alpha_{i0}| = O_p\{(T/\log n)^{-1/2}\}$.

Proof of Lemma B.3. By Lemmas B.2 and C.2, both $n^{-1}\sum_{i=1}^n \boldsymbol\gamma_i I_{ni}^{(1)}$ and $I_n^{(2)}$ are of order $o_p\{\max_{1\le i\le n}|\hat\alpha_i - \alpha_{i0}| + \|\hat{\boldsymbol\beta} - \boldsymbol\beta_0\|\}$. Thus, by (B.5),

$$
\begin{aligned}
\hat{\boldsymbol\beta} - \boldsymbol\beta_0 = &\Gamma_n^{-1}\{-n^{-1}\textstyle\sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0)\boldsymbol\gamma_i + \mathbb{H}_n^{(2)}(\boldsymbol\alpha_0, \boldsymbol\beta_0)\} \\
&+ o_p\{\max_{1\le i\le n}|\hat\alpha_i - \alpha_{i0}| + \|\hat{\boldsymbol\beta} - \boldsymbol\beta_0\|\},
\end{aligned}
$$

which implies that

$$
\begin{aligned}
\|\hat{\boldsymbol\beta} - \boldsymbol\beta_0\| &= O_p\{(nT)^{-1/2} + h_n^r\} + o_p\{\max_{1\le i\le n}|\hat\alpha_i - \alpha_{i0}|\} \\
&= o_p\{T^{-1/2} + \max_{1\le i\le n}|\hat\alpha_i - \alpha_{i0}|\}.
\end{aligned}
$$

Using this estimate, together with Lemma B.2, we obtain by (B.4)

$$
\hat\alpha_i - \alpha_{i0} = s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0) + \bar o_p\{T^{-1/2} + \max_{1\le i\le n}|\hat\alpha_i - \alpha_{i0}|\}.
$$

Now, $\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0) = \{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0) - H_{ni}(\alpha_{i0}, \boldsymbol\beta_0)\} + H_{ni}(\alpha_{i0}, \boldsymbol\beta_0) = \{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0) - H_{ni}(\alpha_{i0}, \boldsymbol\beta_0)\} + \bar O(h_n^r)$ and, by Hoeffding's inequality, it is shown that

$$
\max_{1\le i\le n}|\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0) - H_{ni}(\alpha_{i0}, \boldsymbol\beta_0)| = O_p\{(T/\log n)^{-1/2}\}.
$$

Therefore, we complete the proof. □

By Lemmas B.1-B.3,

$$
\begin{aligned}
\hat\alpha_i - \alpha_{i0} &= s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0) - \boldsymbol\gamma_i'(\hat{\boldsymbol\beta} - \boldsymbol\beta_0) - 2^{-1}s_i f_i^{(1)}(0)(\hat\alpha_i - \alpha_{i0})^2 \\
&\quad - s_i\{(\hat f_i - \mathrm{E}[\hat f_i]) - h_n(\tilde f_i^{(1)} - \mathrm{E}[\tilde f_i^{(1)}])\}(\hat\alpha_i - \alpha_{i0}) + \bar o_p(T^{-1} + \|\hat{\boldsymbol\beta} - \boldsymbol\beta_0\|) \\
&= s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol\beta_0) + \bar O_p\{\|\hat{\boldsymbol\beta} - \boldsymbol\beta_0\| + (\log n)^{3/2}/(Th_n^{1/2})\}, \qquad \text{(B.12)}
\end{aligned}
$$

where we have used Lemma C.2 to obtain the last equality. Put

$$
\begin{aligned}
B_1 &:= n^{-1}\textstyle\sum_{i=1}^n s_i\boldsymbol{\gamma}_i\{(\hat{f}_i - \mathrm{E}[\hat{f}_i]) - h_n(\tilde{f}_i^{(1)} - \mathrm{E}[\tilde{f}_i^{(1)}])\}\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0),\\
B_2 &:= n^{-1}\textstyle\sum_{i=1}^n s_i\{(\hat{\boldsymbol{g}}_i - \mathrm{E}[\hat{\boldsymbol{g}}_i]) - h_n(\tilde{\boldsymbol{g}}_i^{(1)} - \mathrm{E}[\tilde{\boldsymbol{g}}_i^{(1)}])\}\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0),\\
B_3 &:= (2n)^{-1}\textstyle\sum_{i=1}^n s_i^2\boldsymbol{\nu}_i\{\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\}^2.
\end{aligned}
$$

Plugging the expression in (B.12) into (B.9) and (B.10), and using Lemma C.2, we obtain

$$
n^{-1}\textstyle\sum_{i=1}^n \boldsymbol{\gamma}_i I_{in}^{(1)} = -B_1 + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|),\ I_n^{(2)} = -B_2 + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|). \quad \text{(B.13)}
$$

On the other hand, since $\max_{1\le i\le n}|\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)| = O_p\{h_n^r + (T/\log n)^{-1/2}\}$,

$$
n^{-1}\textstyle\sum_{i=1}^n \boldsymbol{\nu}_i(\hat{\alpha}_i - \alpha_{i0})^2 = B_3 + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|). \quad \text{(B.14)}
$$

Plugging the estimates in (B.13)-(B.14) into (B.5), we obtain

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \Gamma_n^{-1}\{&-n^{-1}\textstyle\sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\}\\
&+ \Gamma_n^{-1}(B_1 - B_2 + B_3) + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|). \quad \text{(B.15)}
\end{aligned}
$$

The first term on the right hand side of (B.15) is the sum of independent random variables with approximately zero mean, and expected to follow the central limit theorem. We will establish this in Lemma B.4. On the other hand, B_1, B_2 and B_3 are expected to contribute to the bias. We will establish the limiting behavior of these terms in Lemmas B.5 and B.6.

Lemma B.4. *We have*

$$
\sqrt{nT}\{-n^{-1}\textstyle\sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\} \xrightarrow{d} N\{\mathbf{0}, \tau(1-\tau)V\}.
$$

Proof of Lemma B.4. For an arbitrarily fixed $\boldsymbol{c} \in \mathbb{R}^p$, put $z_{it} := \{\tau - G_{h_n}(u_{it}) + h_n\tilde{K}_{h_n}(u_{it})\}\boldsymbol{c}'(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)$. Note that z_{it} depends on n. By the Cramér-Wold device, it suffices to show that $(nT)^{-1/2}\sum_{i=1}^n\sum_{t=1}^T z_{it} \xrightarrow{d} N\{0, \tau(1-\tau)\boldsymbol{c}'V\boldsymbol{c}\}$. Observe that $(nT)^{-1/2}\sum_{i=1}^n\sum_{t=1}^T z_{it} = (nT)^{-1/2}\sum_{i=1}^n\sum_{t=1}^T (z_{it} - \mathrm{E}[z_{it}]) + O\{(nT)^{1/2}h_n^r\}$ by Lemma C.1 and the second term is $o(1)$ because of the hypothesis of the theorem. A simple algebra yields that

$$
\mathrm{E}[z_{it}^2] = \mathrm{E}[\{\{\tau - G_{h_n}(u_{it})\}\boldsymbol{c}'(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)\}^2] + \bar{O}(h_n).
$$

Now, uniformly over both $\boldsymbol{x}$ and i,

$$
\begin{aligned}
&\mathrm{E}[\{\tau - G_{h_n}(u_{it})\}^2 | \boldsymbol{x}_{it} = \boldsymbol{x}] \\
&= \tau^2 - 2\tau \mathrm{E}[G_{h_n}(u_{it}) | \boldsymbol{x}_{it} = \boldsymbol{x}] + \mathrm{E}[G_{h_n}^2(u_{it}) | \boldsymbol{x}_{it} = \boldsymbol{x}] \\
&= -\tau^2 + O(h_n^r) + \int_{-\infty}^{\infty} \int_{u/h_n}^{\infty} \int_{u/h_n}^{\infty} K(v_1) K(v_2) dv_1 dv_2 f_i(u|\boldsymbol{x}) du \\
&= -\tau^2 + O(h_n^r) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i(h_n \min\{v_1, v_2\} | \boldsymbol{x}) K(v_1) K(v_2) dv_1 dv_2 \\
&= \tau(1-\tau) + O(h_n).
\end{aligned} \tag{B.16}
$$

Thus, the variance of $(nT)^{-1/2} \sum_{i=1}^n \sum_{t=1}^T (z_{it} - \mathrm{E}[z_{it}])$ is $\tau(1-\tau)\boldsymbol{c}'V\boldsymbol{c} + o(1)$. Applying the Lyapunov central limit theorem, we obtain the desired result. □

Lemma B.5. *B_3 is expanded as*

$$
B_3 = \frac{\tau(1-\tau)}{2T} \cdot \left(\frac{1}{n} \sum_{i=1}^n s_i^2 \boldsymbol{\nu}_i \right) + O_p(T^{-1} h_n).
$$

Proof of Lemma B.5. Put $z_{it} := \tau - G_{h_n}(u_{it}) + h_n \tilde{K}_{h_n}(u_{it})$. Then, by Lemma C.1 and (B.16),

$$
\begin{aligned}
\mathrm{E}[\{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\}^2] &= \mathrm{E}[\{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - \mathrm{E}[\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)]\}^2] + (\mathrm{E}[\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)])^2 \\
&= T^{-1} \operatorname{Var}(z_{it}) + \bar{O}(h_n^{2r}) \\
&= T^{-1} \mathrm{E}[z_{it}^2] + \bar{O}(h_n^{2r}) \\
&= T^{-1} \tau(1-\tau) + \bar{O}(T^{-1} h_n + h_n^{2r}).
\end{aligned}
$$

Thus, $\mathrm{E}[B_3] = \tau(1-\tau)(n^{-1} \sum_{i=1}^n s_i^2 \boldsymbol{\nu}_i)/(2T) + O(T^{-1} h_n)$. For an arbitrarily fixed $\boldsymbol{c} \in \mathbb{R}^p$, put $a_i := s_i^2 \boldsymbol{c}' \boldsymbol{\nu}_i$. Then,

$$
\begin{aligned}
&\operatorname{Var}\{n^{-1} \textstyle\sum_{i=1}^n a_i (T^{-1} \sum_{t=1}^T z_{it})^2\} \\
&= n^{-2} \textstyle\sum_{i=1}^n a_i^2 \operatorname{Var}\{(T^{-1} \sum_{t=1}^T z_{it})^2\} \\
&\le n^{-2} \textstyle\sum_{i=1}^n a_i^2 \mathrm{E}[\{(T^{-1} \sum_{t=1}^T z_{it})^4] \\
&= n^{-2} \textstyle\sum_{i=1}^n a_i^2 T^{-4} \sum_{k=1}^T \sum_{l=1}^T \sum_{s=1}^T \sum_{t=1}^T \mathrm{E}[z_{ik} z_{il} z_{is} z_{it}].
\end{aligned} \tag{B.17}
$$

Since $\mathrm{E}[z_{it}] = \bar{O}(h_n^r)$ and $|z_{it}| \le L$ for some constant L depending only on $K(\cdot)$,

$$
\mathrm{E}[z_{ik} z_{il} z_{is} z_{it}] \le \begin{cases} L^4 & k = l = s = t \text{ or } k = l \ne s = t, \\ \bar{O}(h_n^r) & k \ne l = s = t, \\ \bar{O}(h_n^{2r}) & k = l \text{ and } l, s, t \text{ are distinct}, \\ \bar{O}(h_n^{4r}) & k, l, s, t \text{ are distinct}. \end{cases}
$$

Thus, (B.17) is shown to be $O\{n^{-1}(T^{-2} + T^{-2} h_n^r + T^{-1} h_n^{2r} + h_n^{4r})\} = O(n^{-1} T^{-2})$, which leads to the desired result. □

Lemma B.6. *B_1 and B_2 are expanded as*

$$B_1 = \frac{2\tau - 1}{2T} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\gamma}_i \right) + O_p(T^{-1}h_n), \; B_2 = \frac{2\tau - 1}{2T} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\gamma}_i \right) + O_p(T^{-1}h_n).$$

Proof of Lemma B.6. We only prove the expansion for B_1. The expansion for B_2 can be shown in a similar way. Put $z_{it} := \tau - G_{h_n}(u_{it}) + h_n \tilde{K}_{h_n}(u_{it})$. Recall that $\mathrm{E}[z_{it}] = \bar{O}(h_n^r)$ by Lemma C.1. Observe that

$$\begin{aligned}
\mathrm{E}[(\hat{f}_i - \mathrm{E}[\hat{f}_i])(T^{-1} \textstyle\sum_{t=1}^{T} z_{it})] &= \mathrm{E}[(\hat{f}_i - \mathrm{E}[\hat{f}_i]) z_{it}] = T^{-1}\mathrm{E}[\{K_{h_n}(u_{it}) - \mathrm{E}[K_{h_n}(u_{it})]\} z_{it}] \\
&= T^{-1}\mathrm{E}[K_{h_n}(u_{it}) z_{it}] + \bar{O}(T^{-1}h_n^r),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}[K_{h_n}(u_{it}) z_{it}] &= \mathrm{E}[K_{h_n}(u_{it})\{\tau - G_{h_n}(u_{it})\}] + h_n \mathrm{E}[K_{h_n}(u_{it}) \tilde{K}_{h_n}(u_{it})] \\
&= \int_{-\infty}^{\infty} K(u)\{\tau - G(u)\} f_i(uh_n) du + \int_{-\infty}^{\infty} uK^2(u) f_i(uh_n) du \\
&= f_i(0) \left\{ \tau - \int_{-\infty}^{\infty} K(u)G(u) du + \int_{-\infty}^{\infty} uK^2(u) du \right\} + \bar{O}(h_n).
\end{aligned}$$

Since $K(\cdot)$ is symmetric about the origin, the third term inside the brace is zero, and

$$\int_{-\infty}^{\infty} K(u)G(u) du = \left[-\frac{1}{2} G^2(u) \right]_{-\infty}^{\infty} = \frac{1}{2}.$$

Thus, we have $\mathrm{E}[K_{h_n}(u_{it}) z_{it}] = (\tau - 1/2) f_i(0) + \bar{O}(h_n)$. Similarly, observe that

$$\mathrm{E}[(\tilde{f}_i^{(1)} - \mathrm{E}[\tilde{f}_i^{(1)}])(T^{-1} \textstyle\sum_{t=1}^{T} z_{it})] = -T^{-1} h_n^{-2} \mathrm{E}[\tilde{K}^{(1)}(u_{it}/h_n) z_{it}] + \bar{O}(T^{-1}h_n^r),$$

and

$$\begin{aligned}
&h_n^{-2}\mathrm{E}[\tilde{K}^{(1)}(u_{it}/h_n) z_{it}] \\
&= h_n^{-2}\mathrm{E}[\tilde{K}^{(1)}(u_{it}/h_n)\{\tau - G_{h_n}(u_{it})\}] + h_n^{-1}\mathrm{E}[\tilde{K}^{(1)}(u_{it}/h_n) \tilde{K}_{h_n}(u_{it})] \\
&= h_n^{-1} \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u)\{\tau - G(u)\} f_i(uh_n) du + h_n^{-1} \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u) \tilde{K}(u) f_i(uh_n) du \\
&= -h_n^{-1} \int_{-\infty}^{\infty} \tilde{K}(u) K(u) f_i(uh_n) du + \bar{O}(1) \\
&= -h_n^{-1} f_i(0) \int_{-\infty}^{\infty} uK^2(u) du + \bar{O}(1) = \bar{O}(1).
\end{aligned}$$

The third equality is due to the fact that

$$\begin{aligned}
&\int_{-\infty}^{\infty} \tilde{K}^{(1)}(u) \tilde{K}(u) f_i(uh_n) du \\
&= -\int_{-\infty}^{\infty} \tilde{K}(u) \tilde{K}^{(1)}(u) f_i(uh_n) du - h_n \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u) \tilde{K}(u) f_i^{(1)}(uh_n) du \\
&= -\int_{-\infty}^{\infty} \tilde{K}^{(1)}(u) \tilde{K}(u) f_i(uh_n) du + \bar{O}(h_n).
\end{aligned}$$

Therefore, we have $\mathrm{E}[B_1] = (2\tau - 1) \cdot (n^{-1} \sum_{i=1}^n \boldsymbol{\gamma}_i)/(2T) + O(T^{-1} h_n)$.

For an arbitrarily fixed $\boldsymbol{c} \in \mathbb{R}^p$, put $a_i := s_i \boldsymbol{c}' \boldsymbol{\gamma}_i$. Then,

$$
\begin{aligned}
&\mathrm{Var}\{n^{-1} \textstyle\sum_{i=1}^n a_i (\hat{f}_i - \mathrm{E}[\hat{f}_i])(T^{-1} \sum_{t=1}^T z_{it})\} \\
&\quad = n^{-2} \textstyle\sum_{i=1}^n a_i^2 \,\mathrm{Var}\{(\hat{f}_i - \mathrm{E}[\hat{f}_i])(T^{-1} \sum_{t=1}^T z_{it})\} \\
&\quad \le n^{-2} \textstyle\sum_{i=1}^n a_i^2 \mathrm{E}[(\hat{f}_i - \mathrm{E}[\hat{f}_i])^2 (T^{-1} \sum_{t=1}^T z_{it})^2] \\
&\quad \le 2n^{-2} \textstyle\sum_{i=1}^n a_i^2 \mathrm{E}[(\hat{f}_i - \mathrm{E}[\hat{f}_i])^2 \{T^{-1} \sum_{t=1}^T (z_{it} - \mathrm{E}[z_{it}])\}^2] \\
&\qquad + 2n^{-2} \textstyle\sum_{i=1}^n a_i^2 (\mathrm{E}[z_{i1}])^2 \mathrm{E}[(\hat{f}_i - \mathrm{E}[\hat{f}_i])^2],
\end{aligned} \tag{B.18}
$$

and the second term on the right hand side of (B.18) is $O(n^{-1}T^{-1}h_n^{2r-1})$. Put $w_{it} := K_{h_n}(u_{it})$, $\bar{w}_{it} := w_{it} - \mathrm{E}[w_{i1}]$ and $\bar{z}_{it} := z_{it} - \mathrm{E}[z_{i1}]$. Observe that

$$
\mathrm{E}[(\hat{f}_i - \mathrm{E}[\hat{f}_i])^2 \{T^{-1} \textstyle\sum_{t=1}^T (z_{it} - \mathrm{E}[z_{it}])\}^2] = T^{-4} \sum_{k=1}^T \sum_{l=1}^T \sum_{s=1}^T \sum_{t=1}^T \mathrm{E}[\bar{w}_{ik} \bar{w}_{il} \bar{z}_{is} \bar{z}_{it}]
$$

and

$$
\mathrm{E}[\bar{w}_{ik} \bar{w}_{il} \bar{z}_{is} \bar{z}_{it}] = \begin{cases} \mathrm{E}[\bar{w}_{i1}^2 \bar{z}_{i1}^2] & k = l = s = t, \\ \mathrm{E}[\bar{w}_{i1}^2] \mathrm{E}[\bar{z}_{i1}^2] & k = l \neq s = t, \\ (\mathrm{E}[\bar{w}_{i1} \bar{z}_{i1}])^2 & k = s \neq l = t \text{ or } k = t \neq l = s, \\ 0 & \text{otherwise.} \end{cases}
$$

Recall that $|z_{it}| \le L$ for some constant L depending only on $K(\cdot)$, so

$$
\begin{aligned}
&\mathrm{E}[\bar{w}_{i1}^2 \bar{z}_{i1}^2] \le 4L^2 \mathrm{E}[\bar{w}_{i1}^2] = \bar{O}(h_n^{-1}), \\
&\mathrm{E}[\bar{w}_{i1}^2] \mathrm{E}[\bar{z}_{i1}^2] \le 4L^2 \mathrm{E}[\bar{w}_{i1}^2] = \bar{O}(h_n^{-1}), \text{ and} \\
&(\mathrm{E}[\bar{w}_{i1} \bar{z}_{i1}])^2 \le 4L^2 (\mathrm{E}[|\bar{w}_{i1}|])^2 = \bar{O}(1).
\end{aligned}
$$

Thus, the first term on the right hand side of (B.18) is $O(n^{-1}T^{-2}h_n^{-1})$. Similarly, we can show that $\mathrm{Var}\{n^{-1} \sum_{i=1}^n a_i (\tilde{f}_i^{(1)} - \mathrm{E}[\tilde{f}_i^{(1)}])(T^{-1} \sum_{t=1}^T z_{it})\}$ is $O(n^{-1}T^{-2}h_n^{-3})$. Therefore, we have $B_1 = \mathrm{E}[B_1] + O_p\{(n^{-1}T^{-2}h_n^{-1})^{1/2}\}$, which gives the desired result (recall that $n^{-1/2}h_n^{-1/2} = o(h_n)$). □

By (B.15) and Lemmas B.4-B.6, $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\{(nT)^{-1/2} + T^{-1}\} = O_p\{(nT)^{-1/2}\}$, and

$$
\sqrt{nT} \cdot \Gamma_n^{-1}(B_1 - B_2 + B_3) = \sqrt{\frac{n}{T}} \Gamma_n^{-1} \left[\frac{\tau(1-\tau)}{2} \cdot \left(\frac{1}{n} \sum_{i=1}^n s_i^2 \boldsymbol{\nu}_i \right) \right] + O_p(h_n) \xrightarrow{p} \sqrt{\rho} \boldsymbol{b},
$$

where $\boldsymbol{b}$ is given in (3.2). Therefore, we obtain (3.1). □

B.3 Proof of Theorem 3.2

It suffices to show that $\hat{\boldsymbol{b}} \xrightarrow{p} \boldsymbol{b}$. To do this, we shall show that

$$\max_{1\le i\le n} |\hat{s}_i - s_i| \xrightarrow{p} 0, \ \max_{1\le i\le n} \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\| \xrightarrow{p} 0, \ \hat{\Gamma}_n \xrightarrow{p} \Gamma. \tag{B.19}$$

We use the notation in Appendix C below. Without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. Then, $\hat{s}_i$ can be written as $\hat{s}_i = 1/\hat{f}_i(\hat{\alpha}_i, \hat{\boldsymbol{\beta}})$, where $\hat{f}_i(\alpha, \boldsymbol{\beta})$ stands for $\hat{f}_i^{(0)}(\alpha, \boldsymbol{\beta})$. By Lemma C.2, we have $\hat{f}_i(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) = \mathrm{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})]|_{\alpha=\hat{\alpha}_i, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \bar{o}_p(1)$. Observe that $\mathrm{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})] = \{\mathrm{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})] - \mathrm{E}[\hat{f}_i(0, \mathbf{0})]\} + \mathrm{E}[\hat{f}_i(0, \mathbf{0})]$, $\mathrm{E}[\hat{f}_i(0, \mathbf{0})] = f_i(0) + \bar{o}(1)$ and

$$\begin{aligned}
&|\mathrm{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})] - \mathrm{E}[\hat{f}_i(0, \mathbf{0})]| \\
&\le \mathrm{E}\left[\int_{-\infty}^{\infty} |K(u)\{f_i(uh_n + \alpha + \boldsymbol{x}_{it}'\boldsymbol{\beta}|\boldsymbol{x}_{it}) - f_i(uh_n|\boldsymbol{x}_{it})\}|du\right] \\
&\le A_0(|\alpha| + M\|\boldsymbol{\beta}\|)\int_{-\infty}^{\infty} |K(u)|du.
\end{aligned} \tag{B.20}$$

By Proposition 3.1, we have $\hat{f}_i(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) = f_i(0) + \bar{o}_p(1)$, which, by condition (A5) (c), implies that $\hat{s}_i = s_i + \bar{o}_p(1)$. Similarly, we can show that $\hat{\boldsymbol{\gamma}}_i = \boldsymbol{\gamma}_i + \bar{o}_p(1)$. To prove the second assertion of (B.19), observe that

$$\begin{aligned}
\hat{\boldsymbol{\nu}}_i &= -\hat{\boldsymbol{g}}_i^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\gamma}}_i \hat{f}_i^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) \\
&= -\mathrm{E}[\hat{\boldsymbol{g}}_i^{(1)}(\alpha, \boldsymbol{\beta})]|_{\alpha=\hat{\alpha}_i, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \bar{o}_p(1) + \{\boldsymbol{\gamma}_i + \bar{o}_p(1)\}\{\mathrm{E}[\hat{f}_i^{(1)}(\alpha, \boldsymbol{\beta})]|_{\alpha=\hat{\alpha}_i, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \bar{o}_p(1)\} \\
&= -\mathrm{E}[\hat{\boldsymbol{g}}_i^{(1)}(0, \mathbf{0})] + \bar{o}_p(1) + \{\boldsymbol{\gamma}_i + \bar{o}_p(1)\}\{\mathrm{E}[\hat{f}_i^{(1)}(0, \mathbf{0})] + \bar{o}_p(1)\} \\
&= \boldsymbol{\nu}_i + \bar{o}_p(1),
\end{aligned}$$

where the second equality is due to Lemma C.2. Third inequality can be deduced from the evaluation analogous to (B.20). Similarly, we can show that $\hat{\Gamma}_n = \Gamma_n + o_p(1)$ and by condition (A8) (a), $\hat{\Gamma}_n^{-1} = \Gamma_n^{-1} + o_p(1)$.

Finally, since s_i and $\boldsymbol{\nu}_i$ are bounded over $i \ge 1$, we have

$$\begin{aligned}
\hat{\boldsymbol{b}} &= \{\Gamma_n^{-1} + o_p(1)\}\left[\frac{\tau(1-\tau)}{2} \cdot \frac{1}{n}\sum_{i=1}^{n}\{s_i^2 + \bar{o}_p(1)\}\{\boldsymbol{\nu}_i + \bar{o}_p(1)\}\right] \\
&= \Gamma_n^{-1}\left[\frac{\tau(1-\tau)}{2} \cdot \frac{1}{n}\sum_{i=1}^{n} s_i^2 \boldsymbol{\nu}_i\right] + o_p(1) \xrightarrow{p} \boldsymbol{b}.
\end{aligned}$$

Therefore, we complete the proof.

□

C Miscellaneous lemmas

In this section, we summarize some miscellaneous results used in the proof of Theorem 3.1. Lemma C.2 is a modification of Lemma 3 in Horowitz (1998). Throughout the section, we assume the conditions of Theorem 3.1.

Lemma C.1. *We have*

(a) $H_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) = \bar{O}(h_n^r),\ H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = O(h_n^r)$;

(b) $\partial_{\alpha_i} H_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) = -f_i(0) + \bar{O}(h_n^{r-1})$;

(c) $\partial_{\alpha_i}^2 H_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) = -f_i^{(1)}(0) + \bar{O}(h_n^{r-2})$;

(d) $\partial_{\boldsymbol{\beta}} H_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) = \partial_{\alpha_i} H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = -\mathrm{E}[f_i(0|\boldsymbol{x}_{it})\boldsymbol{x}_{it}] + \bar{O}(h_n^{r-1})$;

(e) $\partial_{\boldsymbol{\beta}'} H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = -n^{-1}\sum_{i=1}^n \mathrm{E}[f_i(0|\boldsymbol{x}_{it})\boldsymbol{x}_{it}\boldsymbol{x}_{it}'] + O(h_n^{r-1})$;

(f) $\partial_{\alpha_i}^2 H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = -\mathrm{E}[f_i^{(1)}(0|\boldsymbol{x}_{it})\boldsymbol{x}_{it}] + \bar{O}(h_n^{r-2})$;

(g) $\partial_{\alpha_i}\partial_{\beta_j} H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}),\ \partial_{\beta_j}\partial_{\beta_k} H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}),\ \partial_{\alpha_i}^3 H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}),\ \partial_{\alpha_i}^2\partial_{\beta_j} H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}),\ \partial_{\alpha_i}\partial_{\beta_j}\partial_{\beta_k} H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta})$ *and* $\partial_{\beta_j}\partial_{\beta_k}\partial_{\beta_l} H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta})$ *are* $O(1)$ *uniformly over both* $(\alpha_i, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{B}$ *and* $1 \leq i \leq n$ *for* $1 \leq j, k, l \leq p$*, that is, for instance,* $\max_{1 \leq i \leq n} \sup_{(\alpha_i, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{B}} \partial_{\alpha_i}\partial_{\beta_j} H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}) = O(1)$ *as* $n \to \infty$ *for* $1 \leq j \leq p$*;*

(h) $\partial_{\alpha_i}\partial_{\beta_j} H_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}),\ \partial_{\beta_j}\partial_{\beta_k} H_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}),\ \partial_{\alpha_i}^3 H_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \partial_{\alpha_i}^2\partial_{\beta_j} H_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}),\ \partial_{\alpha_i}\partial_{\beta_j}\partial_{\beta_k} H_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *and* $\partial_{\beta_j}\partial_{\beta_k}\partial_{\beta_l} H_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *are* $O(1)$ *uniformly over both* $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A}^n \times \mathcal{B}$ *and* $1 \leq i \leq n$ *for* $1 \leq j, k, l \leq p$.

Proof. The proof is immediate from conditions (A2), (A5) and (A6). □

Put

$$
\begin{aligned}
&\hat{f}_i^{(j)}(\alpha, \boldsymbol{\beta}) := (-1)^j (Th_n^{j+1})^{-1} \textstyle\sum_{t=1}^T K^{(j)}((u_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})/h_n),\ j = 0, 1,\\
&\tilde{f}_i^{(j)}(\alpha, \boldsymbol{\beta}) := (-1)^j (Th_n^{j+1})^{-1} \textstyle\sum_{t=1}^T \tilde{K}^{(j)}((u_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})/h_n),\ j = 1, 2,\\
&\hat{\boldsymbol{g}}_i^{(j)}(\alpha, \boldsymbol{\beta}) := (-1)^j (Th_n^{j+1})^{-1} \textstyle\sum_{t=1}^T K^{(j)}((u_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})/h_n)\boldsymbol{x}_{it},\ j = 0, 1,\\
&\tilde{\boldsymbol{g}}_i^{(j)}(\alpha, \boldsymbol{\beta}) := (-1)^j (Th_n^{j+1})^{-1} \textstyle\sum_{t=1}^T \tilde{K}^{(j)}((u_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})/h_n)\boldsymbol{x}_{it},\ j = 1, 2,\\
&\hat{J}_i(\alpha, \boldsymbol{\beta}) := (Th_n)^{-1} \textstyle\sum_{t=1}^T K((u_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})/h_n)\boldsymbol{x}_{it}\boldsymbol{x}_{it}',\\
&\tilde{J}_i^{(1)}(\alpha, \boldsymbol{\beta}) := -(Th_n^2)^{-1} \textstyle\sum_{t=1}^T \tilde{K}^{(1)}((u_{it} - \alpha - \boldsymbol{x}_{it}'\boldsymbol{\beta})/h_n)\boldsymbol{x}_{it}\boldsymbol{x}_{it}',
\end{aligned}
$$

where $K^{(j)}(u) = d^j K(u)/du^j$ and $K^{(0)}(u)$ stands for $K(u)$. The same rule applies to $\tilde{K}(u)$.

Lemma C.2. *Uniformly over both* $(\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}$ *and* $1 \leq i \leq n$*,*

(a) $\hat{f}_i^{(j)}(\alpha, \boldsymbol{\beta}) - \mathrm{E}[\hat{f}_i^{(j)}(\alpha, \boldsymbol{\beta})] = o_p\{\log n/(Th_n^{2j+1})^{1/2}\}, \ j = 0, 1;$

(b) $\tilde{f}_i^{(j)}(\alpha, \boldsymbol{\beta}) - \mathrm{E}[\tilde{f}_i^{(j)}(\alpha, \boldsymbol{\beta})] = o_p\{\log n/(Th_n^{2j+1})^{1/2}\}, \ j = 1, 2;$

(c) $\hat{\boldsymbol{g}}_i^{(j)}(\alpha, \boldsymbol{\beta}) - \mathrm{E}[\hat{\boldsymbol{g}}_i^{(j)}(\alpha, \boldsymbol{\beta})] = o_p\{\log n/(Th_n^{2j+1})^{1/2}\}, \ j = 0, 1;$

(d) $\tilde{\boldsymbol{g}}_i^{(j)}(\alpha, \boldsymbol{\beta}) - \mathrm{E}[\tilde{\boldsymbol{g}}_i^{(j)}(\alpha, \boldsymbol{\beta})] = o_p\{\log n/(Th_n^{2j+1})^{1/2}\}, \ j = 1, 2;$

(e) $\hat{J}_i(\alpha, \boldsymbol{\beta}) - \mathrm{E}[\hat{J}_i(\alpha, \boldsymbol{\beta})] = o_p\{\log n/(Th_n)^{1/2}\};$

(f) $\tilde{J}_i^{(1)}(\alpha, \boldsymbol{\beta}) - \mathrm{E}[\tilde{J}_i^{(1)}(\alpha, \boldsymbol{\beta})] = o_p\{\log n/(Th_n^3)^{1/2}\}.$

As in Horowitz (1998), we employ empirical process theory to prove the lemma. Before the proof, we introduce some notation. 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 $\|\cdot\|_{L_2(Q)}$. In the proof of Lemma C.2, we make use of an exponential inequality for centered empirical processes, which originates from Talagrand (1996). The current form of the inequality is due to Corollary 2.2 in Gine and Guillou (2002). For a reference, we describe the inequality in the next proposition. There, 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 $\mathcal{G}$ 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 $\xi \in S$.[11] This condition is discussed in Section 2.3 of van der Vaart and Wellner (1996).

Proposition C.1 (Talagrand (1996); in this form, Gine and Guillou (2002))**.** *Let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$. Let $\mathcal{F}$ be a pointwise measurable class of functions on $(S, \mathcal{S})$ uniformly bounded by a constant U and such that for some constants $A \geq 3\sqrt{e}$ and $v \geq 1$, $N(\mathcal{F}, L_2(Q), U\epsilon) \leq (A/\epsilon)^v$ for $0 < \epsilon < 1$ and for every probability measure Q on $(S, \mathcal{S})$. Assume that $\mathrm{E}[f(\xi_1)] = 0$ for all $f \in \mathcal{F}$. Let σ^2 be such that $\sigma^2 \geq \sup_{f \in \mathcal{F}} \mathrm{E}[f^2(\xi_1)]$. If, moreover, $0 < \sigma < U/2$ and $\sqrt{n}\sigma \geq U\sqrt{\log(U/\sigma)}$, there exists positive constants L and C depending only on A and v such that for all t satisfying*

$$C\sqrt{n}\sigma\sqrt{\log\frac{U}{\sigma}} \leq t \leq C\frac{n\sigma^2}{U},$$

we have

$$\mathrm{P}\left\{\left\|\sum_{i=1}^n f(\xi_i)\right\|_{\mathcal{F}} > t\right\} \leq L\exp\left\{-\frac{\varphi(L, C)t^2}{n\sigma^2}\right\},$$

where $\varphi(L, C) := \log(1 + C/(4L))/(LC)$.

[11] The assumption that $\mathcal{F}$ is pointwise measurable is substantial for Proposition 3.1 since the proof of Talagrand's (1996) Theorem 1.4 exploits the fact that the class is countable.

Now, we shall prove Lemma C.2.

Proof of Lemma C.2. We only prove (a) with $j = 0$. The other cases can be shown in a similar way. For simplicity, we write $\hat{f}_i(\alpha, \boldsymbol{\beta})$ for $\hat{f}_i^{(0)}(\alpha, \boldsymbol{\beta})$. Put $\xi_{it} := (u_{it}, \boldsymbol{x}_{it})$, $g_{\alpha,\boldsymbol{\beta},h}(u, \boldsymbol{x}) := K((u - \alpha - \boldsymbol{x}'\boldsymbol{\beta})/h)$ for $(\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}$ and $h > 0$, $\mathcal{G}_n := \{g_{\alpha,\boldsymbol{\beta},h_n} : (\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}\}$ and $\tilde{\mathcal{G}}_{ni} := \{g - \mathrm{E}[g(\xi_{i1})] : g \in \mathcal{G}_n\}$. It suffices to show that for every $\epsilon > 0$,

$$\mathrm{P}\left(\max_{1 \le i \le n} \left\| \sum_{t=1}^{T} g(\xi_{it}) \right\|_{\tilde{\mathcal{G}}_{ni}} > \epsilon \sqrt{T h_n} \log n \right) \to 0.$$

Because of the subadditivity of the probability measure, the left hand side is bounded by

$$\sum_{i=1}^{n} \mathrm{P}\left(\left\| \sum_{t=1}^{T} g(\xi_{it}) \right\|_{\tilde{\mathcal{G}}_{ni}} > \epsilon \sqrt{T h_n} \log n \right),$$

so it suffices to show that

$$\max_{1 \le i \le n} \mathrm{P}\left(\left\| \sum_{t=1}^{T} g(\xi_{it}) \right\|_{\tilde{\mathcal{G}}_{ni}} > \epsilon \sqrt{T h_n} \log n \right) = o(n^{-1}).$$

We apply Proposition C.1 to $\tilde{\mathcal{G}}_{ni}$. To do this, we shall check the conditions of Proposition C.1 to $\tilde{\mathcal{G}}_{ni}$. The continuity of $K(\cdot)$ guarantees that $\tilde{\mathcal{G}}_{ni}$ is pointwise measurable. Put $\kappa := \sup_{u \in \mathbb{R}} |K(u)|$. We can take $U = 2\kappa$ to $\tilde{\mathcal{G}}_{ni}$. Next, we check the metric entropy condition. Put $\mathcal{G} := \{g_{\alpha,\boldsymbol{\beta},h} : (\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}, h > 0\}$ and $\tilde{\mathcal{G}}_i := \{g - \mathrm{E}[g(\xi_{i1})] : g \in \mathcal{G}\}$. Since condition (A6) guarantees that $K(\cdot)$ is of bounded variation, by Lemma 22 of Nolan and Pollard (1987), there exist constants A and v such that $N(\mathcal{G}, L_2(Q), \kappa\delta) \le (A/\delta)^v$ for $0 < \delta < 1$ and every probability measure Q on $\mathbb{R}^{p+1}$. Let P_i denote the joint distribution of $(u_{it}, \boldsymbol{x}_{it})$. Since $|\mathrm{E}[g(\xi_{i1})] - \mathrm{E}[\bar{g}(\xi_{i1})]| \le \mathrm{E}[|g(\xi_{i1}) - \bar{g}(\xi_{i1})|] \le \|g - \bar{g}\|_{L_2(P_i)}$ for $g, \bar{g} \in \mathcal{G}$, if we write $\mathcal{G}_{\delta,Q}$ for a δ-cover of $\mathcal{G}$ with respect to $\|\cdot\|_{L_2(Q)}$, the class $\{g - \mathrm{E}[\bar{g}(\xi_{i1})] : g \in \mathcal{G}_{\delta,Q}, \bar{g} \in \mathcal{G}_{\delta,P_i}\}$ is a (2δ)-cover of $\tilde{\mathcal{G}}_i$ with respect to $\|\cdot\|_{L_2(Q)}$. Thus, for every $0 < \delta < 1$ and every probability measure Q on $\mathbb{R}^{p+1}$, we have

$$\begin{aligned} N(\tilde{\mathcal{G}}_{ni}, L_2(Q), 2\kappa\delta) &\le N(\tilde{\mathcal{G}}_i, L_2(Q), 2\kappa\delta) \\ &\le N(\mathcal{G}, L_2(Q), \kappa\delta) N(\mathcal{G}, L_2(P_i), \kappa\delta) \\ &\le (A/\delta)^{2v}. \end{aligned}$$

Let L, C be the constants given in Proposition C.1. Observe that

$$\begin{aligned} \mathrm{E}[g^2_{\alpha,\boldsymbol{\beta},h_n}(\xi_{i1})] &= h_n \mathrm{E}\left[\int_{-\infty}^{\infty} K^2(u) f_i(u h_n + \alpha + \boldsymbol{x}'_{i1}\boldsymbol{\beta} | \boldsymbol{x}_{i1}) du \right] \\ &\le h_n A_0 \int_{-\infty}^{\infty} K^2(u) du =: \sigma_n^2, \end{aligned}$$

for $(\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}$ and $i \geq 1$. Finally, since $h_n \asymp T^{-a}$ with $1/r < a < 1/3$ and $T \asymp n$, there exists a positive integer n_0 (independent of i) such that for all $n \geq n_0$,

$$0 < \sigma_n < \kappa, \ \sqrt{T}\sigma_n \geq 2\kappa\sqrt{\log(2\kappa/\sigma_n)}, \ C\sqrt{T}\sigma_n\sqrt{\log\frac{2\kappa}{\sigma_n}} \leq \epsilon\sqrt{Th_n}\log n \leq C\frac{T\sigma_n^2}{2\kappa}.$$

Therefore, by Proposition C.1, we have for $n \geq n_0$,

$$\mathrm{P}\left(\left\|\sum_{t=1}^{T} g(\xi_{it})\right\|_{\tilde{\mathcal{G}}_{ni}} > \epsilon\sqrt{Th_n}\log n\right) \leq L\exp\left\{-\frac{\varphi(L,C)\epsilon^2(\log n)^2}{A_0\|K\|_{L_2(du)}^2}\right\},$$

which is $o(n^{-1})$ and the right hand side is independent of i. The proof ends. □

References

Abrevaya, J., and C. M. Dahl (2008): "The Effects of Birth Inputs on Birthweight: Evidence From Quantile Estimation on Panel Data," *Journal of Business and Economic Statistics*, 26, 379–397.

Amemiya, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

Angrist, J., V. Chernozhukov, and I. Fernandez-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. 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.

Arellano, M., and S. Bonhomme (2009): "Robust Priors in Nonlinear Panel Data Models," *Econometrica*, 77, 489–536.

Arellano, M., and J. Hahn (2005): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," Mimeo, CEMFI.

Bester, C. A., and C. Hansen (2009): "A Penalty Function Approach to Bias Reduction in Non-Linear Panel Models with Fixed Effects," *Journal of Business and Economic Statistics*, 27, 131–148.

Buchinsky, M. (1995): "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics*, 68, 303–338.

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

Dhaene, G., and K. Jochmans (2009): "Split-Panel Jacknife Estimation of Fixed-Effect Models," Mimeo.

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

Gine, E., and A. Guillou (2002): "Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators," *Annales de l'Institut Henri Poincare*, 38, 907–921.

Graham, B. S., J. Hahn, and J. L. Powell (2009): "The Incidental Parameter Problem in a Non-Differentiable Panel Data Model," *Economics Letters*, 105, 181–182.

Hahn, J., and G. M. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When both n and T are Large," *Econometrica*, 70, 1639–1657.

——— (2004): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When both n and T are Large," Mimeo.

Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319.

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

——— (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66, 1327–1351.

——— (2001): "The Bootstrap in Econometrics," in *Handbook of Econometrics, Vol.5*, ed. by J. J. Heckman, and E. E. Leamer, pp. 3159–3228. Elsevier Science B.V.

Ichimura, H., and P. Todd (2007): "Implementing Nonparametric and Semiparametric Estimators," in *Handbook of Econometrics, Vol.6*, ed. by J. J. Heckman, and E. E. Leamer, pp. 5369–5468. Elsevier Science B.V.

Kato, K., A. F. Galvao, Jr., and G. V. Montes-Rojas (2010): "Asymptotics and Bootstrap Inference for Panel Quantile Regression Models with Fixed Effects," Mimeo.

Knight, K. (1998): "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *Annals of Statistics*, 26, 755–770.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

——— (2005): *Quantile Regression*. Cambridge University Press, Cambridge.

——— (2009): "Quantile Regression in R: A Vignette," Available from: http://cran.r-project.org/web/packages/quantreg/index.html.

Koenker, R., and G. W. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–49.

Lancaster, T. (2000): "The Incidental Parameter Problem since 1948," *Journal of Econometrics*, 95, 391–413.

Li, H., B. G. Lindsay, and R. P. Waterman (2003): "Efficiency of Projected Score Methods in Rectangular Array Asymptotics," *Journal of the Royal Statistical Society, Series B*, 65, 191–208.

Muller, H. G. (1984): "Smooth Optimal Kernel Estimators of Densities, Regression Curves and Modes," *Annals of Statistics*, 12, 766–774.

Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

Nolan, D., and D. Pollard (1987): "U-processes: Rates of Convergence," *Annals of Statistics*, 15, 780–799.

Phillips, P. C., and D. Sul (2007): "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence," *Journal of Econometrics*, 137, 162–188.

Rilstone, P., V. Srivastava, and A. Ullah (1996): "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators," *Journal of Econometrics*, 75, 369–395.

Rosen, A. (2009): "Set Identification via Quantile Restrictions in Short Panels," CEMMAP working paper CWP26/09.

Talagrand, M. (1996): "New Concentration Inequalities in Product Spaces," *Inventiones Mathematicae*, 126, 505–563.

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

Woutersen, T. M. (2002): "Robustness Against Incidental Parameters," Mimeo, University of Western Ontario.

Table 1: Bias and standard deviation. Normal error. $n = 100$.

	T	$T\times$Bias				Stdev.			
	T	$\hat{\beta}_{\text{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$	$\hat{\beta}_{\text{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$
$\tau = 0.25$	8	0.0848	0.0657	0.0288	0.0864	0.0454	0.0452	0.0511	0.0455
	12	0.0981	0.0761	0.0302	0.1010	0.0372	0.0371	0.0429	0.0374
	16	0.1258	0.1020	0.0476	0.1308	0.0318	0.0316	0.0347	0.0319
	20	0.1171	0.0918	0.0122	0.1228	0.0277	0.0276	0.0315	0.0278
$\tau = 0.5$	8	-0.0211	-0.0236	-0.0319	-0.0175	0.0408	0.0406	0.0459	0.0408
	12	-0.0128	-0.0118	-0.0262	-0.0087	0.0342	0.0340	0.0389	0.0343
	16	-0.0021	0.0014	-0.0055	0.0026	0.0287	0.0287	0.0336	0.0288
	20	0.0034	0.0088	-0.0016	0.0077	0.0253	0.0253	0.0262	0.0254
$\tau = 0.75$	8	-0.1196	-0.1163	-0.0962	-0.1157	0.0456	0.0455	0.0513	0.0457
	12	-0.1148	-0.1081	-0.0615	-0.1110	0.0373	0.0372	0.0457	0.0374
	16	-0.1021	-0.0938	-0.0347	-0.0993	0.0318	0.0317	0.0369	0.0318
	20	-0.1119	-0.1027	-0.0299	-0.1094	0.0279	0.0279	0.0306	0.0279

Table 2: Bias and standard deviation. χ^2_3 error. $n = 100$.

	T	$T\times$Bias				Stdev.			
	T	$\hat{\beta}_{\text{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$	$\hat{\beta}_{\text{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$
$\tau = 0.25$	8	0.0098	0.0197	-0.1208	0.0120	0.0617	0.0613	0.0605	0.0618
	12	-0.0157	0.0001	-0.1427	-0.0143	0.0492	0.0491	0.0486	0.0494
	16	-0.0184	0.0001	-0.1434	-0.0164	0.0419	0.0418	0.0414	0.0420
	20	-0.0126	0.0063	-0.1352	-0.0104	0.0373	0.0373	0.0372	0.0375
$\tau = 0.5$	8	-0.1361	-0.1226	-0.2186	-0.1350	0.0868	0.0867	0.0910	0.0871
	12	-0.1769	-0.1608	-0.1938	-0.1793	0.0708	0.0707	0.0735	0.0709
	16	-0.1051	-0.0877	-0.0807	-0.1052	0.0607	0.0607	0.0624	0.0609
	20	-0.1378	-0.1193	-0.0899	-0.1338	0.0547	0.0547	0.0558	0.0549
$\tau = 0.75$	8	-0.4836	-0.4655	-0.3535	-0.4737	0.1364	0.1362	0.1552	0.1363
	12	-0.5368	-0.5121	-0.2122	-0.5253	0.1111	0.1110	0.1330	0.1111
	16	-0.5321	-0.5034	-0.1310	-0.5193	0.0919	0.0918	0.1112	0.0918
	20	-0.4782	-0.4470	0.0180	-0.4534	0.0852	0.0852	0.0943	0.0852

Table 3: Bias and standard deviation. Normal error. $n = 200$.

		$T\times$Bias				Stdev.			
	T	$\hat{\beta}_{\mathrm{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$	$\hat{\beta}_{\mathrm{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$
$\tau = 0.25$	8	0.0974	0.0834	0.0428	0.0887	0.0328	0.0324	0.0363	0.0325
	12	0.1011	0.0916	0.0353	0.0995	0.0261	0.0259	0.0306	0.0260
	16	0.1090	0.1012	0.0382	0.1110	0.0228	0.0227	0.0262	0.0229
	20	0.0965	0.0907	0.0152	0.1023	0.0195	0.0194	0.0206	0.0196
$\tau = 0.5$	8	-0.0008	-0.0016	-0.0048	-0.0005	0.0296	0.0296	0.0336	0.0297
	12	0.0058	0.0060	-0.0031	0.0065	0.0238	0.0238	0.0267	0.0238
	16	0.0022	0.0033	-0.0030	0.0022	0.0206	0.0206	0.0222	0.0206
	20	-0.0043	-0.0024	-0.0118	-0.0041	0.0182	0.0182	0.0198	0.0183
$\tau = 0.75$	8	-0.1006	-0.0996	-0.0842	-0.1000	0.0332	0.0332	0.0543	0.0332
	12	-0.0816	-0.0796	-0.0308	-0.0808	0.0267	0.0267	0.0307	0.0268
	16	-0.0895	-0.0870	-0.0251	-0.0890	0.0224	0.0224	0.0248	0.0224
	20	-0.1096	-0.1065	-0.0332	-0.1091	0.0197	0.0197	0.0211	0.0197

Table 4: Bias and standard deviation. χ^2_3 error. $n = 200$.

		$T\times$Bias				Stdev.			
	T	$\hat{\beta}_{\mathrm{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$	$\hat{\beta}_{\mathrm{KB}}$	$\hat{\beta}$	$\hat{\beta}^1$	$\hat{\beta}_{1/2}$
$\tau = 0.25$	8	0.0121	0.0143	-0.1259	0.0119	0.0433	0.0433	0.0429	0.0433
	12	-0.0074	-0.0028	-0.1479	-0.0077	0.0340	0.0340	0.0337	0.0340
	16	-0.0215	-0.0157	-0.1593	-0.0206	0.0296	0.0296	0.0294	0.0296
	20	-0.0007	0.0053	-0.1350	-0.0010	0.0260	0.0260	0.0259	0.0261
$\tau = 0.5$	8	-0.1344	-0.1305	-0.2312	-0.1338	0.0629	0.0628	0.0650	0.0629
	12	-0.1415	-0.1356	-0.1725	-0.1401	0.0484	0.0484	0.0506	0.0485
	16	-0.1706	-0.1643	-0.1577	-0.1693	0.0417	0.0418	0.0427	0.0418
	20	-0.1359	-0.1291	-0.1007	-0.1345	0.0372	0.0372	0.0380	0.0373
$\tau = 0.75$	8	-0.4706	-0.4658	-0.3446	-0.4691	0.0967	0.0967	0.1118	0.0968
	12	-0.5038	-0.4971	-0.2028	-0.5016	0.0750	0.0750	0.0895	0.0750
	16	-0.5464	-0.5379	-0.1363	-0.5432	0.0674	0.0674	0.0776	0.0673
	20	-0.5500	-0.5415	-0.0781	-0.5482	0.0588	0.0588	0.0651	0.0588