

Smoothed quantile regression for panel data*

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

September 21, 2011

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 panel data models. 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, and provide the analytic form of the asymptotic bias. We propose a one-step bias correction to the fixed effects estimator based on the analytic bias formula obtained from our asymptotic analysis. Importantly, our results cover the case that the observations are dependent in the time dimension. 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.

*The authors are grateful to the seminar participants at Kyoto University, 16th International Conference on Panel Data, New York Camp Econometrics V, and 2011 North American Winter and Summer Meetings of the Econometric Society. Kato's work is supported by the Grant-in-Aid for Scientific research (KAKENHI 22730179) from the JSPS.

[†]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 Iowa and 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 effects 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.

Recently, panel data sets have become widely available and very popular, because they provide a large number of data points and allow researchers to study dynamics of adjustment as well as to control for individual specific heterogeneity. In this regard, it is natural to consider to apply QR for panel data. In fact, there is a rapidly expanding empirical literature on QR for panel data in important areas of economics and biostatistics, which makes a persuasive case for the rigorous study of estimation and inference properties of the QR method for panel data.¹ In contrast to mean regression, however, little is known about the ability of the QR method to analyze panel models with individual effects. 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 altogether (Koenker, 2004).² Unfortunately, the resulting estimator of the common parameters will be generally 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 when T

¹For example, in applied economics, Abrevaya and Dahl (2008), Silva, Kosmopoulou, and Lamarche (2009), and Gamper-Rabindran, Khan, and Timmins (2010); in biostatistics, Lipsitz, Fitzmaurice, Molenberghs, and Zhao (1997), Wei and He (2006) and Wang and Fygenon (2009), among others.

²Koenker (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).

grows at 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), Hahn and Kuersteiner (2011), 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 case since it substantially depends on the smoothness of objective functions. Kato, Galvao, and Montes-Rojas (2011) formally established the asymptotic properties of the standard fixed effects 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 standard QR estimator in panel models is partly explained by the fact that the higher order stochastic expansion of the scores is an essential technical tool in the asymptotic analysis of Hahn and Newey (2004) and Hahn and Kuersteiner (2011) while such an expansion is difficult to implement to the QR objective function because the Taylor series method can not be applied to it. It is also important to note that the higher order asymptotic behavior of QR estimators is non-standard and rather complicated (Arcones, 1998; Knight, 1998). See also the discussion at the end of Section 5 of Kato, Galvao, and Montes-Rojas (2011). To our knowledge, there is no paper that formally studies the bias correction for panel QR models under the large n and T asymptotics.

This paper overcomes the above problem by smoothing the objective function. The idea of smoothing non-differentiable objective functions goes back 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 its limiting normal distribution has a bias in the mean (even) when n and T grow at the same rate. Importantly, we allow the observations to be dependent in the time dimension. As naturally expected, the bias depends on the conditional and unconditional densities and their first derivatives. We propose a 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 included in the previous literature 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 the present 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. An additional complication arises since the observations are dependent in the time dimension. To cope with this complication, we derive new empirical process inequalities.

We conduct a small Monte Carlo study to illustrate the effect of the bias correction to the FE-SQR estimator. The results show that, 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 introduce a panel 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

We consider a panel QR model with individual effects

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

where $\tau \in (0, 1)$ is a quantile of interest, y_{it} is a dependent variable, \mathbf{x}_{it} is a p dimensional vector of explanatory variables, α_{i0} is a scalar individual effect for each i and $Q_\tau(y_{it}|\mathbf{x}_{it}, \alpha_{i0})$ is the conditional τ -quantile of y_{it} given $(\mathbf{x}_{it}, \alpha_{i0})$. In general, each α_{i0} 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.³ The model is semiparametric in the sense that the functional form of the conditional distribution of y_{it} given $(\mathbf{x}_{it}, \alpha_{i0})$ is left unspecified and no parametric assumption is made on the relation between α_{i0} and \mathbf{x}_{it} . The number of individuals is denoted by n and the number of time periods is denoted by $T = T_n$ that may depends on n . In what follows, we omit the subscript n of T_n . It should be noted that although the model (2.1) looks “linear” at a first

³Some readers might wonder that the individual effects should be independent of the quantile. However, in our model, the individual effects include the intercept term and the intercept term should depend on the quantile. Thus, it would be natural that the individual effects depend on the quantile.

sight, the quantile is a nonlinear functional of the random variable, so in general the conventional differencing out strategy is not appropriate for the QR case.

We consider the fixed effects estimation of β_0 , which is implemented by treating each individual effect as a parameter to be estimated. Throughout the paper, as in Hahn and Newey (2004), Hahn and Kuersteiner (2011) and Fernandez-Val (2005), we treat α_{i0} as fixed by conditioning on them. Then, the standard QR estimation solves

$$\min_{\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^p} \left[\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \alpha_i - \mathbf{x}'_{it}\beta) \{\tau - I(y_{it} \leq \alpha_i + \mathbf{x}'_{it}\beta)\} \right], \quad (2.2)$$

where $I(\cdot)$ is the indicator function. Note that α implicitly depends on n . As in general nonlinear panel models, except for some special cases, the estimator of β_0 defined by (2.2) is inconsistent as $n \rightarrow \infty$ while T is fixed. In fact, as Graham, Hahn, and Powell (2009) noted, when $T = 2$, the standard fixed effects QR estimator of the common parameter does not depend on τ for panel quantile regression models with individual effects; however since the common parameter depends on τ unless the model is of a linear location form that Graham, Hahn, and Powell (2009) presumed, so the standard fixed effects QR estimator will be generally inconsistent.⁴

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 (2011) rigorously investigated the asymptotic properties of the standard QR estimator, say $\hat{\beta}_{\text{KB}}$, defined by (2.2) when n and T jointly go to infinity. The essential content of their theoretical results is that while $\hat{\beta}_{\text{KB}}$ is consistent under mild regularity conditions, justifying (mean-zero) asymptotic normality of $\hat{\beta}_{\text{KB}}$ requires the restrictive condition that T grows faster than n^2 .⁵ They remarked that the non-smoothness of the objective function (more precisely, the non-smoothness of the scores) significantly complicates the asymptotic analysis of $\hat{\beta}_{\text{KB}}$ 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 $\hat{\beta}_{\text{KB}}$ is not known even for the case that the observations are independent in both cross section and time dimensions.

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

⁴A more explicit example in which the standard fixed effects QR estimator is inconsistent is available upon request.

⁵Kato, Galvao, and Montes-Rojas (2011), however, did not show that this condition is necessary for (mean-zero) asymptotic normality of $\hat{\beta}_{\text{KB}}$. Whether one can substantially improve this condition is still an open question.

objective function.⁶ Smoothing the QR objective function is employed in Horowitz (1998) to study the bootstrap refinement for inference in conditional quantile models.⁷ The basic insight of Horowitz (1998) is to smooth over $I(y_{it} \leq \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta})$ by using a kernel function. Let $K(\cdot)$ be a kernel function and $G(\cdot)$ be the survival function of $K(\cdot)$, i.e.,

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

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 $h_n \rightarrow 0$ as $n \rightarrow \infty$ and write $G_{h_n}(\cdot) = G(\cdot/h_n)$. Note that $G_{h_n}(y_{it} - \alpha_i - \mathbf{x}'_{it}\boldsymbol{\beta})$ is a smoothed counterpart of $I(y_{it} \leq \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta})$. Then, we consider the estimator

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

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 the set $\mathcal{A}^n \times \mathcal{B}$ 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 $\partial_{\alpha_i} = \partial/\partial\alpha_i$ 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

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

where

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

⁶In fact, it turns out that using smoothing is crucial to develop the formal asymptotic analysis here, since, as Hahn and Newey (2004) and Hahn and Kuersteiner (2011) observed, in a general nonlinear panel model with a smooth objective function, the bias of the fixed effects estimator of the common parameter depends on the second order biases of the estimators of the individual parameters. However, as Kato, Galvao, and Montes-Rojas (2011) discussed, the second order bias of the general QR estimator (in the cross section case) is not uniquely determined, so there is a conceptual difficulty in deriving a bias result to the unsmoothed fixed effects QR estimator.

⁷A motivation to use 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).

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 $\hat{\beta}$. This can be understood by regarding $\tilde{K}(\cdot)$ as a “kernel”, although 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 β as the β -derivative. $n^{-1}\mathbb{H}_{ni}^{(1)}(\alpha_i, \beta)$ and $\mathbb{H}_n^{(2)}(\alpha, \beta)$ are the α_i -derivative and the β -derivative of one minus the objective function in (2.3), respectively. From an analogy to the ML estimation, we call $\mathbb{H}_{ni}^{(1)}(\alpha_i, \beta)$ the α_i -score and $\mathbb{H}_n^{(2)}(\alpha, \beta)$ the β -score.

3 Main results

In this section, we investigate the asymptotic properties of the FE-SQR estimator defined by (2.3). Put $\alpha_0 := (\alpha_{10}, \dots, \alpha_{n0})'$.

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 $n/T \rightarrow 0$. However, we believe that the conditions are, to some extent, standard in the literature.

(A1) For each $i \geq 1$, the process $\{(y_{it}, \mathbf{x}_{it}), t = 0, \pm 1, \pm 2, \dots\}$ is stationary and β -mixing. Let $\beta_i(j)$ denote the β -mixing coefficient of the process $\{(y_{it}, \mathbf{x}_{it}), t = 0, \pm 1, \pm 2, \dots\}$.⁸ Assume that there exist constants $a \in (0, 1)$ and $B \geq 0$ such that $\sup_{i \geq 1} \beta_i(j) \leq Ba^j$. Furthermore, the processes $\{(y_{it}, \mathbf{x}_{it}), t = 0, \pm 1, \pm 2, \dots\}$ are independent across i .

(A2) There exists a constant M such that $\sup_{i \geq 1} \|\mathbf{x}_{i1}\| \leq M$ (a.s.).

(A3) The parameter sets $\mathcal{A} \subset \mathbb{R}$ and $\mathcal{B} \subset \mathbb{R}^p$ are compact. For each $i \geq 1$, α_{i0} is an interior point of \mathcal{A} , and β_0 is an interior point of \mathcal{B} .

⁸For any stationary process $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$, the β -mixing coefficient is defined as

$$\beta(k) := \frac{1}{2} \sup \left\{ \sum_{i=1}^I \sum_{j=1}^J |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| : \{A_i\}_{i=1}^I \text{ is any finite partition in } \mathcal{A}_{-\infty}^0, \right. \\ \left. \{B_j\}_{j=1}^J \text{ is any finite partition in } \mathcal{A}_k^\infty \right\},$$

where \mathcal{A}_i^j is the σ -field generated by X_i, \dots, X_j for $i < j$.

The distribution of $(y_{it}, \mathbf{x}_{it})$ is allowed to depend on i . Define $u_{it} := y_{it} - \alpha_{i0} - \mathbf{x}'_{it}\boldsymbol{\beta}_0$. Condition (A1) implies that the process $\{(u_{it}, \mathbf{x}_{it}), t = 0, \pm 1, \pm 2, \dots\}$ is stationary and β -mixing for each i with β -mixing coefficient $\beta_i(\cdot)$, and independent across i . Let $F_i(u|\mathbf{x})$ denote the conditional distribution function of u_{it} given $\mathbf{x}_{it} = \mathbf{x}$. We assume that $F_i(u|\mathbf{x})$ has a density $f_i(u|\mathbf{x})$. Let $f_i(u)$ denote the marginal density of u_{it} . To cope with the time series dependence, we further assume that for each $i \geq 1$ and $j = \pm 1, \pm 2, \dots$, there exists a joint conditional density of $(u_{i1}, u_{i,1+j})$ given $(\mathbf{x}_{i1}, \mathbf{x}_{i,1+j}) = (\mathbf{x}_1, \mathbf{x}_{1+j})$, denoted as $f_{i,j}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j})$. Let $f_{i,j}(u_1, u_{1+j})$ denote the joint (unconditional) density of $(u_{i1}, u_{i,1+j})$. 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 $E[f_i(0|\mathbf{x}_{i1})(1, \mathbf{x}'_{i1})'(1, \mathbf{x}'_{i1})]$ are bounded away from zero uniformly over $i \geq 1$.

(A5) (a) For each $i \geq 1$, $f_i(u|\mathbf{x})$ is r times continuously differentiable with respect to u for each fixed \mathbf{x} , where $r \geq 4$ is an even integer. Let $f_i^{(j)}(u|\mathbf{x}) := (\partial/\partial u)^j f_i(u|\mathbf{x})$ for $j = 0, 1, \dots, r$, where $f_i^{(0)}(u|\mathbf{x})$ stands for $f_i(u|\mathbf{x})$. (b) There exists a constant C_f such that $|f_i^{(j)}(u|\mathbf{x})| \leq C_f$ uniformly over (u, \mathbf{x}) for all $j = 0, 1, \dots, r$ and $i \geq 1$. (c) $\inf_{i \geq 1} f_i(0) > 0$. (d) For each $i \geq 1$ and $j = \pm 1, \pm 2, \dots$, $f_{i,j}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j})$ is r times continuously differentiable with respect to (u_1, u_{1+j}) for each fixed $(\mathbf{x}_1, \mathbf{x}_{1+j})$. Let

$$f_{i,j}^{(k,l)}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j}) := \partial^{k+l} f_{i,j}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j}) / \partial u_1^k \partial u_{1+j}^l,$$

where $f_{i,j}^{(0,0)}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j})$ stands for $f_{i,j}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j})$. (e) There exists a constant C'_f such that $|f_{i,j}^{(k,0)}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j})| \leq C'_f$ uniformly over $(u_1, u_{1+j}, \mathbf{x}_1, \mathbf{x}_{1+j})$ for all $0 \leq k \leq r, i \geq 1$ and $j = \pm 1, \pm 2, \dots$. (f) There exists an integrable function $D_f : \mathbb{R} \rightarrow [0, \infty)$ such that $|f_{i,j}^{(0,l)}(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j})| \leq D_f(u_{1+j})$ uniformly over $(u_1, u_{1+j}, \mathbf{x}_1, \mathbf{x}_{1+j})$ for all $0 \leq l \leq r, i \geq 1$ and $j = \pm 1, \pm 2, \dots$.

(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, i.e.,

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

where r is given in condition (A5).

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

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

(A8) (a) $\Gamma_n := n^{-1} \sum_{i=1}^n \mathbb{E}[f_i(0|\mathbf{x}_{i1})\mathbf{x}_{i1}(\mathbf{x}'_{i1} - \boldsymbol{\gamma}'_i)]$ is nonsingular for each n , and the limit $\Gamma := \lim_{n \rightarrow \infty} \Gamma_n$ exists and is nonsingular. (b) Let V_{ni} denote the covariance matrix of the term $T^{-1/2} \sum_{t=1}^T \{\tau - I(u_{it} \leq 0)\}(\mathbf{x}_{it} - \boldsymbol{\gamma}_i)$. Assume the limit $V := \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n V_{ni}$ exists and is nonsingular.

(A9) Define

$$\begin{aligned}\omega_{ni}^{(1)} &:= \sum_{1 \leq |j| \leq T-1} \left(1 - \frac{|j|}{T}\right) \left\{ \tau f_i(0) - \int_{-\infty}^0 f_{i,j}(0, u) du \right\}, \\ \omega_{ni}^{(2)} &:= \sum_{1 \leq |j| \leq T-1} \left(1 - \frac{|j|}{T}\right) \left\{ \tau \mathbb{E}[f_i(0|\mathbf{x}_{i1})\mathbf{x}_{i1}] - \mathbb{E} \left[\mathbf{x}_{i1} \int_{-\infty}^0 f_{i,j}(0, u|\mathbf{x}_{i1}, \mathbf{x}_{i,1+j}) du \right] \right\}, \\ \omega_{ni}^{(3)} &:= \sum_{|j| \leq T-1} \left(1 - \frac{|j|}{T}\right) \text{Cov}\{I(u_{i1} \leq 0), I(u_{i,1+j} \leq 0)\}.\end{aligned}$$

Assume that $\max_{1 \leq i \leq n} |\omega_{ni}^{(1)}| = O(1)$, $\max_{1 \leq i \leq n} \|\omega_{ni}^{(2)}\| = (1)$, and the limit $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n s_i(\omega_{ni}^{(1)} \boldsymbol{\gamma}_i - \omega_{ni}^{(2)} + s_i \omega_{ni}^{(3)} \boldsymbol{\nu}_i/2)$ exists.

Condition (A1) is the same as Condition 1 of Hahn and Kuersteiner (2011) except for that we now assume β -mixing instead of α -mixing. Condition (A1) allows for the following two possibilities:

- (a) All regressors are strictly exogenous and some of these exogenous regressors are generated by a time series model.
- (b) Lagged dependent variables are included as explanatory variables.

The reason to assume a β -mixing condition instead of α -mixing is to establish convergence rates of kernel type statistics. For that purpose, β -mixing is more tractable than α -mixing. We refer to Section 2.6 of Fan and Yao (2003) for some basic materials on mixing processes. The condition that \mathbf{x}_{it} is uniformly bounded can be relaxed at the expense of lengthier proofs. Condition (A3) is a conventional assumption in the literature. As Kato, Galvao, and Montes-Rojas (2011) noted, condition (A4) guarantees identification of each α_{i0} and $\boldsymbol{\beta}_0$. Conditions (A5) (a)-(c) state some restrictions on the conditional densities, and are standard in the QR literature (see Assumption 4 of Horowitz (1998) and condition (ii) of Angrist, Chernozhukov, and Fernandez-Val (2006, Theorem 3)). We require no less than four times differentiability of the conditional densities to make the smoothing bias negligible. Conditions (A5) (d)-(f) are mainly used to derive an explicit bias expression (see the proof of Lemma A.6).

Condition (A6) corresponds to Assumption 5 of 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 $c < 1/3$ is explained in Horowitz

(1998). The proof of Theorem 3.1 below shows that the asymptotic bias of $\hat{\beta}$ 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 $cr > 1$. A higher order kernel is needed to ensure condition (A7). Several higher order kernels are introduced 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. The exact form of the term V_{ni} is given by

$$V_{ni} = \sum_{|j| \leq T-1} \left(1 - \frac{|j|}{T}\right) \mathbb{E}[\{\tau - I(u_{i1} \leq 0)\}\{\tau - I(u_{i,1+j} \leq 0)\}(\mathbf{x}_{i1} - \boldsymbol{\gamma}_i)(\mathbf{x}_{i,1+j} - \boldsymbol{\gamma}_i)'].$$

If there is no time series dependence, i.e., for each i , the process $\{(y_{it}, \mathbf{x}_{it}), t = 0, \pm 1, \pm 2, \dots\}$ is i.i.d., then $V_{ni} = \tau(1 - \tau)\mathbb{E}[(\mathbf{x}_{i1} - \boldsymbol{\gamma}_i)(\mathbf{x}_{i1} - \boldsymbol{\gamma}_i)']$, $\omega_{ni}^{(1)} = 0$, $\omega_{ni}^{(2)} = \mathbf{0}$ and $\omega_{ni}^{(3)} = \tau(1 - \tau)$.

3.2 Limiting distribution

We first show consistency of the estimator. Although the main concern is the distributional result, the 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 (2011), 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 given in Appendix A.1.

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

We now 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 given in Appendix A.2.

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

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

where

$$\mathbf{b} := \Gamma^{-1} \left[\lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{i=1}^n s_i \left(\omega_{ni}^{(1)} \boldsymbol{\gamma}_i - \omega_{ni}^{(2)} + \frac{s_i \omega_{ni}^{(3)} \boldsymbol{\nu}_i}{2} \right) \right\} \right]. \quad (3.2)$$

The bias (3.2) is of a complicated form. In the next subsection, to gain some insight into the bias expression, we compare our bias formula with the Hahn-Newey formula

in the simplest case where the observations are independent in the time dimension. By the theorem, one can see that the bias consists of three terms. As the proof shows, the first one comes from the correlation between $\hat{\alpha}_i$ and the α_i -derivative of the α_i -score evaluated at the truth; the second one comes from the correlation between $\hat{\alpha}_i$ and the α_i -derivative of the β -score evaluated at the truth; the third one comes from the variance of $\hat{\alpha}_i$. See the expression (A.16) and Lemmas A.5-A.6 in Appendix A.

Our proof strategy of Theorem 3.1 is substantially different from a functional expansion method of Li, Lindsay, and Waterman (2003) that many papers exploited. The reason is that their technique is difficult to adapt to the situation where the kernel smoothing is used. 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 present objective function is smooth, 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. An additional complication arises since the observations are dependent in the time dimension. To cope with this complication, we derive new empirical process inequalities applicable to β -mixing processes, which are presented in Appendix C.

3.3 A heuristic comparison with the Hahn-Newey formula

In this subsection, we provide a heuristic comparison of our bias formula with the Hahn-Newey formula in the case that the observations are independent in the time dimension.⁹ Hahn and Newey (2004) gave the bias formula for the general fixed effects M -estimator with smooth objective functions in the case that the observations are independent in both the cross section and time dimensions (Hahn and Newey (2004) focused on the likelihood setting, but as they stated, the bias expression presented in Hahn and Newey (2004, p.1302) does not depend on the likelihood setting). Observe that when the observations are independent in the time dimension, the bias term \mathbf{b} reduces to a simpler form

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

In what follows, with a heuristic approach, we present that our bias formula is identical to the Hahn-Newey formula applied to the (unsmoothed) QR objective function. In

⁹It is possible to compare our result with the bias expression derived in Hahn and Kuersteiner (2011) for the case that the observations are dependent in the time dimension. However, the purpose of this section is to provide an intuition behind our bias formula, and to make the argument simpler, we restrict our attention to the case that the observations are independent in the time dimension.

the present context, “heuristic” means that there is no formal justification in the mechanical application of the Hahn-Newey formula to the QR objective function as demonstrated below.

To begin with, we briefly review the Hahn-Newey formula. Let $f(\mathbf{z}_{it}; \alpha, \boldsymbol{\beta})$ denote some smooth objective function to be maximized, where \mathbf{z}_{it} is a random vector of conformal dimension, α is a scalar constant and $\boldsymbol{\beta}$ is a constant vector of conformal dimension. Let α_{i0} and $\boldsymbol{\beta}_0$ denote the true parameter values. Define

$$v_{it}(\alpha, \boldsymbol{\beta}) := \frac{\partial}{\partial \alpha} f(\mathbf{z}_{it}; \alpha, \boldsymbol{\beta}), \quad \mathbf{w}_{it}(\alpha, \boldsymbol{\beta}) := \frac{\partial}{\partial \boldsymbol{\beta}} f(\mathbf{z}_{it}; \alpha, \boldsymbol{\beta}).$$

As in Hahn and Newey (2004), we use the notation $v_{it} = v_{it}(\alpha_{i0}, \boldsymbol{\beta}_0)$ and $\mathbf{w}_{it} = \mathbf{w}_{it}(\alpha_{i0}, \boldsymbol{\beta}_0)$. The Hahn-Newey bias formula is given by

$$\text{Bias} = \sqrt{\rho} \times \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathcal{I}_i \right)^{-1} \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{b}_i \right),$$

where $\rho := \lim_{n \rightarrow \infty} (n/T)$ and

$$\begin{aligned} \mathcal{I}_i &:= -\mathbf{E} \left[\mathbf{w}_{it\beta} - \frac{\mathbf{E}[\mathbf{w}_{it\alpha}]}{\mathbf{E}[v_{it\alpha}]} v'_{it\beta} \right], \\ \mathbf{b}_i &:= \left[\frac{\mathbf{E}[\mathbf{w}_{it\alpha}]}{\mathbf{E}[v_{it\alpha}]^2} \mathbf{E}[v_{it\alpha} v_{it}] \right] - \left[\frac{\mathbf{E}[\mathbf{w}_{it\alpha} v_{it}]}{\mathbf{E}[v_{it\alpha}]} \right] + \left[\frac{\mathbf{E}[v_{it}^2]}{2\mathbf{E}[v_{it\alpha}]^2} \left\{ \mathbf{E}[\mathbf{w}_{it\alpha\alpha}] - \frac{\mathbf{E}[\mathbf{w}_{it\alpha}]}{\mathbf{E}[v_{it\alpha}]} \mathbf{E}[v_{it\alpha\alpha}] \right\} \right] \\ &=: \mathbf{b}_{1i} - \mathbf{b}_{2i} + \mathbf{b}_{3i}. \end{aligned}$$

The terms like $v_{it\alpha}$ are defined as $v_{it\alpha} = \partial v_{it}(\alpha, \boldsymbol{\beta}) / \partial \alpha |_{(\alpha, \boldsymbol{\beta}) = (\alpha_{i0}, \boldsymbol{\beta}_0)}$ and so on. Note that we have changed the original order of terms in \mathbf{b}_i to make the comparison more comprehensible. Here, \mathbf{b}_{1i} , \mathbf{b}_{2i} and \mathbf{b}_{3i} correspond to what we have called the sources of the bias in the previous subsection. Also we have slightly changed the original notation to avoid the notational multiplication.

We now turn to the QR case. Assume without loss of generality $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. The objective function is $f(\mathbf{z}_{it}; \alpha, \boldsymbol{\beta}) = -\rho_\tau(u_{it} - \alpha - \mathbf{x}'_{it}\boldsymbol{\beta})$ with $\mathbf{z}_{it} = (u_{it}, \mathbf{x}_{it})$. The scores would be

$$v_{it}(\alpha, \boldsymbol{\beta}) = \tau - I(u_{it} \leq \alpha + \mathbf{x}'_{it}\boldsymbol{\beta}), \quad \mathbf{w}_{it}(\alpha, \boldsymbol{\beta}) = \{\tau - I(u_{it} \leq \alpha + \mathbf{x}'_{it}\boldsymbol{\beta})\} \mathbf{x}_{it}.$$

The scores for the QR objective function are not uniquely determined since the check function $\rho_\tau(u)$ is not differentiable at $u = 0$. However, we shall intuitively calculate the Hahn-Newey bias formula to the present scores based on some heuristic rule. The scores include the indicator function; so we can not compute, for instance, $\mathbf{E}[v_{it\alpha}]$ as it is. Instead, we shall compute $\mathbf{E}[v_{it\alpha}]$ as $\partial \mathbf{E}[v_{it}(\alpha, \boldsymbol{\beta})] / \partial \alpha |_{(\alpha, \boldsymbol{\beta}) = (\alpha_{i0}, \boldsymbol{\beta}_0)}$.¹⁰ Using this

¹⁰Another way to implement this heuristic calculation is to use “generalized functions” as in Phillips (1991).

heuristic rule, we can obtain:

$$\begin{aligned} \mathbb{E}[v_{it\alpha}] &= -f_i(0), \quad \mathbb{E}[v_{it\beta}] = \mathbb{E}[\mathbf{w}_{it\alpha}] = -\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}], \quad \mathbb{E}[v_{it\alpha\alpha}] = -f_i^{(1)}(0), \\ \mathbb{E}[\mathbf{w}_{it\beta}] &= -\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}\mathbf{x}'_{it}], \quad \mathbb{E}[\mathbf{w}_{it\alpha\alpha}] = -\mathbb{E}[f_i^{(1)}(0|\mathbf{x}_{it})\mathbf{x}_{it}]. \end{aligned}$$

The calculation of the terms $\mathbb{E}[v_{it\alpha}v_{it}]$ and $\mathbb{E}[\mathbf{w}_{it\alpha}v_{it}]$ is not straightforward. We make use of the fact that $\mathbb{E}[v_{it\alpha}v_{it}] = \mathbb{E}[\partial v_{it}^2/\partial\alpha]/2$ and $\mathbb{E}[\mathbf{w}_{it\alpha}v_{it}] = \mathbb{E}[v_{it\alpha}v_{it}\mathbf{x}_{it}] = \mathbb{E}[(\partial v_{it}^2/\partial\alpha)\mathbf{x}_{it}]/2$. Exchanging the derivative and the expectation, we can obtain

$$\mathbb{E}[v_{it\alpha}v_{it}] = \frac{(1-2\tau)}{2}f_i(0), \quad \mathbb{E}[\mathbf{w}_{it\alpha}v_{it}] = \frac{(1-2\tau)}{2}\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}].$$

Substituting these results into the Hahn-Newey formula, we can obtain

$$\begin{aligned} \mathcal{I}_i &= \mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}\mathbf{x}'_{it}] - \frac{\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}]}{f_i(0)}\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}'_{it}] = \mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}(\mathbf{x}'_{it} - \gamma'_i)], \\ \mathbf{b}_{1i} &= -\frac{\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}]}{f_i^2(0)} \cdot \frac{(1-2\tau)}{2}f_i(0) = \frac{(2\tau-1)}{2} \cdot \gamma_i, \\ \mathbf{b}_{2i} &= \frac{(2\tau-1)}{2} \cdot \frac{\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}]}{f_i(0)} = \frac{(2\tau-1)}{2} \cdot \gamma_i, \\ \mathbf{b}_{3i} &= \frac{\tau(1-\tau)}{2f_i^2(0)} \left\{ -\mathbb{E}[f_i^{(1)}(0|\mathbf{x}_{it})\mathbf{x}_{it}] + \frac{\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{it}]}{f_i(0)}f_i^{(1)}(0) \right\} = \frac{\tau(1-\tau)}{2} \cdot s_i^2\boldsymbol{\nu}_i. \end{aligned}$$

Therefore, the Hahn-Newey formula reduces to our formula. Note that in the case that the observations are independent in the time dimension, \mathbf{b}_{1i} and \mathbf{b}_{2i} are canceled out, so that the resulting bias is of a simple form.

Of course, the above calculation is not formal because of the non-differentiability of the QR objective function, or more precisely, because the Hahn-Newey bias formula depends on the higher order stochastic expansion of the scores while that for the QR objective function is non-standard, as discussed in Introduction. In the present paper, the smoothing is employed as a device to make the above heuristic calculation rigorous.

3.4 Bias correction

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

We first consider a one-step bias correction based on the analytic form of the asymptotic bias. Put $\hat{u}_{it} := y_{it} - \hat{\alpha}_i - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}}$. The terms $f_i := f_i(0)$, s_i , γ_i , $\boldsymbol{\nu}_i$ and Γ

can be estimated by

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

where $K^{(1)}(u) = dK(u)/du$. The estimation of the terms $\omega_{ni}^{(1)}$, $\omega_{ni}^{(2)}$ and $\omega_{ni}^{(3)}$ is a more delicate issue, since it reduces to the estimation of long run covariances. As in Hahn and Kuersteiner (2011), we make use of a truncation strategy. Define

$$\begin{aligned}\phi_i(j) &:= \int_{-\infty}^0 f_{i,j}(0, u) du, \\ \varphi_i(j) &:= \mathbb{E} \left[\mathbf{x}_{i1} \int_{-\infty}^0 f_{i,j}(0, u | \mathbf{x}_{i1}, \mathbf{x}_{i,1+j}) du \right], \\ \varrho_i(j) &:= \mathbb{E}[I(u_{i1} \leq 0)I(u_{i,1+j} \leq 0)].\end{aligned}$$

Since $\phi_i(j) \approx \mathbb{E}[K_{h_n}(u_{i1})I(u_{i,1+j} \leq 0)]$, it can be estimated by

$$\hat{\phi}_i(j) := \frac{1}{T} \sum_{t=\max\{1, -j+1\}}^{\min\{T, T-j\}} K_{h_n}(\hat{u}_{it}) I(\hat{u}_{i,t+j} \leq 0).$$

Similarly, $\varphi_i(j)$ can be estimated by

$$\hat{\varphi}_i(j) := \frac{1}{T} \sum_{t=\max\{1, -j+1\}}^{\min\{T, T-j\}} K_{h_n}(\hat{u}_{it}) I(\hat{u}_{i,t+j} \leq 0) \mathbf{x}_{it}.$$

The term $\varrho_i(j)$ can be estimated by its sample analogue:

$$\hat{\varrho}_i(j) := \frac{1}{T} \sum_{t=\max\{1, -j+1\}}^{\min\{T, T-j\}} I(\hat{u}_{it} \leq 0) I(\hat{u}_{i,t+j} \leq 0).$$

Take a sequence m_n such that $m_n \rightarrow \infty$ sufficiently slowly. Then, $\omega_{ni}^{(1)}$, $\omega_{ni}^{(2)}$ and $\omega_{ni}^{(3)}$ can be estimated by

$$\begin{aligned}\hat{\omega}_{ni}^{(1)} &:= \sum_{1 \leq |j| \leq m_n} \left(1 - \frac{|j|}{T}\right) \{\tau \hat{f}_i - \hat{\phi}_i(j)\}, \\ \hat{\omega}_{ni}^{(2)} &:= \sum_{1 \leq |j| \leq m_n} \left(1 - \frac{|j|}{T}\right) \{\tau \hat{f}_i \hat{\gamma}_i - \hat{\varphi}_i(j)\}, \\ \hat{\omega}_{ni}^{(3)} &:= \tau(1 - \tau) + \sum_{1 \leq |j| \leq m_n} \left(1 - \frac{|j|}{T}\right) \{-\tau^2 + \hat{\varrho}_i(j)\}.\end{aligned}$$

The bias term \mathbf{b} is thus estimated by

$$\hat{\mathbf{b}} := \hat{\Gamma}_n^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \hat{s}_i \left(\hat{\omega}_{ni}^{(1)} \hat{\gamma}_i - \hat{\omega}_{ni}^{(2)} + \frac{\hat{s}_i \hat{\omega}_{ni}^{(3)} \hat{\nu}_i}{2} \right) \right\}.$$

We define the one-step bias corrected estimator by $\hat{\beta}^1 := \hat{\beta} - \hat{\mathbf{b}}/T$. In practice, there is no need to compute the terms $\tau \hat{f}_i$ and $\tau \hat{f}_i \hat{\gamma}_i$ in $\hat{\omega}_{ni}^{(1)}$ and $\hat{\omega}_{ni}^{(2)}$, respectively, as they are canceled out by the difference $\hat{\omega}_{ni}^{(1)} \hat{\gamma}_i - \hat{\omega}_{ni}^{(2)}$. Additionally, there is no need to use the same kernel and the same bandwidth to estimate β_0 and \mathbf{b} . In particular, the kernel used in the estimation of the bias may not be of a higher order (see the remark after Theorem 3.2 below). However, to make the notation simpler, we do not introduce a new kernel and a new bandwidth.

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

Theorem 3.2. *Let $m_n \rightarrow \infty$ such that $m_n^2 (\log n) / (Th_n^2) \rightarrow 0$. Under the same conditions as Theorem 3.1, we have*

$$\sqrt{nT}(\hat{\beta}^1 - \beta_0) \xrightarrow{d} N(\mathbf{0}, \Gamma^{-1} V \Gamma^{-1}).$$

Remark 3.1. As discussed in Section 3.1, the order of the kernel used to construct the FE-SQR estimator has to be at least 4. However, the order of the kernel used to construct the bias estimator can be 2. In view of the proof of Theorem 3.2, to make $\hat{\mathbf{b}}$ consistent, it is sufficient that h_n^r is of order $o\{(\log n)/(Th_n)^{1/2}\}$, which is satisfied if $h_n \asymp T^{-a}$ with $1/(2r+1) \leq a < 1/3$. In particular, r can be 2 in the estimation of the bias term alone.

Remark 3.2 (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 $\hat{\Gamma}_n \xrightarrow{p} \Gamma$. The matrix V_{ni} can be consistently estimated by

$$\begin{aligned} \hat{V}_{ni} &= \frac{\tau(1-\tau)}{T} \sum_{t=1}^T (\mathbf{x}_{it} - \hat{\gamma}_i)(\mathbf{x}_{it} - \hat{\gamma}_i)' \\ &+ \sum_{1 \leq |j| \leq m_n} \left(1 - \frac{|j|}{T}\right) \left[\frac{1}{T} \sum_{t=\max\{1, -j+1\}}^{\min\{T, T-j\}} \{\tau - I(\hat{u}_{it} \leq 0)\} \{\tau - I(\hat{u}_{i,t+j} \leq 0)\} (\mathbf{x}_{it} - \hat{\gamma}_i)(\mathbf{x}_{i,t+j} - \hat{\gamma}_i)' \right], \end{aligned}$$

so that V_n is consistently estimated by $\hat{V}_n = n^{-1} \sum_{i=1}^n \hat{V}_{ni}$. The proof of the consistency of \hat{V}_n is similar to the proof of the consistency of $\hat{\mathbf{b}}$ given in Appendix A.3. Under the same conditions as Theorem 3.2, it is shown that $\hat{\Gamma}_n^{-1} \hat{V}_n \hat{\Gamma}_n^{-1}$ is consistent for the

asymptotic covariance matrix $\Gamma^{-1}V\Gamma^{-1}$. Here, the term $(1 - |j|/T)$ in the definition of \hat{V}_{ni} can be replaced by 1 or $(1 - |j|/m_n)$, and such a replacement does not affect the consistency property.

Remark 3.3. As discussed in Hahn and Kuersteiner (2011), for small T , a default choice of m_n would be 1. For relatively large T , a choice of m_n could be significant. However, a standard theory of autocorrelation consistent covariance matrix estimation, such as that given in Andrews (1991), can not be applied here because Andrews (1991) restricted moment functions to be smooth while the moment function of quantile regression is not smooth, and the optimal choice of m_n is beyond the scope of the paper.

In the case that the observations are independent in the time dimension, Hahn and Newey (2004, Theorem 2) showed that under suitable regularity conditions, the mean-zero asymptotic normality of the bias corrected MLE for smooth likelihood functions hold when $n/T^3 \rightarrow 0$. In the present case, although we do not exclude the possibility that Theorem 3.2 holds under $n/T^3 \rightarrow 0$, a formal justification will be quite involved and requires even stringent conditions, so that we do not pursue here this direction.

To be precise, consider a fixed effects estimator $\hat{\theta}$ of the common parameter θ_0 for some “standard” panel models (say, panel probit model) with individual effects. If the objective function is smooth enough, and some moment condition is satisfied, then, as Hahn and Newey (2004) and the subsequent study suggested, $\hat{\theta}$ admits the expansion

$$\hat{\theta} - \theta_0 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T U_{it} + \frac{B}{T} + O_p(T^{-2}) + o_p\{(nT)^{-1/2}\}, \quad (3.3)$$

where U_{it} are mean zero random vectors, and B is some constant which eventually contributes to the bias. Importantly, this expansion is valid without $n/T \rightarrow \text{const.}$ (but a suitable growth condition to ensure the consistency of the estimator is required). The condition that $n/T \rightarrow \text{const.}$ comes so that $(nT)^{-1/2}$ and T^{-1} rates are balanced. If we “successfully” remove the bias B , which means that B itself can be estimated with bias of order T^{-1} , then the asymptotic normality (with mean zero) holds under the condition that T^{-2} is faster than $(nT)^{-1/2}$, i.e., $n = o(T^3)$ (this discussion is less formal but explains the main point that Hahn and Newey (2004) did). To summarize, Hahn and Newey’s Theorem 2 crucially depends on the two observations:

- (i) the fixed effects estimator admits the expansion of the form (3.3);
- (ii) the bias term itself can be estimated with bias of order T^{-1} , i.e., $\hat{B} = B + O_p\{T^{-1} + (nT)^{-1/2}\}$.

In contrast, such a “clean” expansion does not hold for the FE-SQR estimator because of the presence of the smoothing bias, and even higher order expansions of the FE-SQR objective function will be quite involved. Additionally, to guarantee an expansion of the form (3.3) in the present case requires considerably stringent conditions. For example, there is a bias term of order h_n^r in the present case, and to make this term of order T^{-2} , r (the order of the kernel) should be at least 8. Implementing (ii) in the present case is also challenging, since the bias now depends on long run covariances. In fact, even for standard smooth objective functions, Hahn and Kuersteiner (2011) only proved the asymptotic normality of the bias corrected estimator under the condition that $n/T \rightarrow \text{const.}$

However, it is still true that the FE-SQR estimator $\hat{\beta}$ admits the expansion of the form (3.3) with the $O_p(T^{-2})$ term replaced by $o_p(T^{-1})$. A careful inspection of the proof shows that this expansion holds under a slightly weaker growth condition on T than $n/T \rightarrow \text{const.}$, although T should not be too slow to ensure sufficient convergence rates of some kernel type statistics (and the consistency of the FE-SQR estimator). Therefore, if we can estimate consistently the bias term \mathbf{b} , the asymptotic normality (with mean zero) holds under a slightly weaker growth condition on T than $n/T \rightarrow \text{const.}$

An undesirable feature of the analytic bias correction is that it depends on the inverse of the estimated densities \hat{f}_i . When T is relatively small, on rare occasions, \hat{f}_i may be very close to zero for some i , which results in an unreasonable value of $\hat{\beta}^1$. To avoid such a situation, a reasonable solution is to use a trimming. Let $\hat{I}_i := I(\hat{f}_i > \kappa_n)$ for some sequence κ_n such that $\kappa_n \rightarrow 0$ slowly. We propose to modify $\hat{\Gamma}_n$ and $\hat{\mathbf{b}}$ as

$$\begin{aligned} \tilde{\Gamma}_n &= \frac{1}{nT} \sum_{i=1}^n \hat{I}_i \sum_{t=1}^T K_{h_n}(\hat{u}_{it}) \mathbf{x}_{it} (\mathbf{x}_{it} - \hat{\gamma}_i), \\ \tilde{\mathbf{b}} &= \tilde{\Gamma}_n^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \hat{I}_i \hat{s}_i \left(\hat{\omega}_{ni}^{(1)} \hat{\gamma}_i - \hat{\omega}_{ni}^{(2)} + \frac{\hat{s}_i \hat{\omega}_{ni}^{(3)} \hat{\nu}_i}{2} \right) \right\}. \end{aligned}$$

The validity of such a modification is simple. By the proof of Theorem 3.2, $\max_{1 \leq i \leq n} |\hat{f}_i - f_i| \xrightarrow{p} 0$, and by condition (A5), $\inf_{i \geq 1} f_i > 0$. Thus, with probability approaching one, $\hat{f}_i > \kappa_n$ for all $1 \leq i \leq n$. This implies that $\hat{\mathbf{b}} = \tilde{\mathbf{b}}$ with probability approaching one. Trimming is frequently used in the nonparametric and semiparametric literatures. We refer to Section 6.4 of Ichimura and Todd (2007) for discussion on trimming.

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 $\hat{\beta}$. 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 $\hat{\beta}_{S_l}$ be the FE-SQR estimate based on the data

$\{(y_{it}, \mathbf{x}_{it}), 1 \leq i \leq n, t \in S_l\}$ for $l = 1, 2$. The half-panel jackknife estimator is defined as $\hat{\beta}_{1/2} := 2\hat{\beta} - \bar{\beta}_{1/2}$, where $\bar{\beta}_{1/2} := (\hat{\beta}_{S_1} + \hat{\beta}_{S_2})/2$. For simplicity, suppose for a moment that we use the same bandwidth to construct $\hat{\beta}$ and $\hat{\beta}_{S_l}$ ($l = 1, 2$). Then, from the asymptotic representation of the FE-SQR estimator that appears in the proof of Theorem 3.1 (see (A.16) in Appendix B), it can be shown that under the same conditions as Theorem 3.1 (including $n/T \rightarrow \rho$), we have $\sqrt{nT}(\hat{\beta}_{1/2} - \beta_0) \xrightarrow{d} N(\mathbf{0}, \Gamma^{-1}V\Gamma^{-1})$. The half-panel jackknife estimator would be attractive in both theoretical and practical senses. In fact, its validity does not require any extra condition. 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 in the literature (see, for instance, Hahn and Newey, 2004; Arellano and Hahn, 2005; Bester and Hansen, 2009; Dhaene and Jochmans, 2009). Although 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.

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:

$$y_{it} = \eta_i + x_{it} + (1 + 0.5x_{it})\epsilon_{it}, \quad (4.1)$$

where $x_{it} = 0.3\eta_i + z_{it}$, $z_{it} \sim \text{i.i.d. } \chi_3^2$, $\eta_i \sim \text{i.i.d. } U[0, 1]$, $\epsilon_{it} = \varepsilon_{it} + \theta\varepsilon_{it-1}$, and $\varepsilon_{it} \sim \text{i.i.d. } F$ with $F = N(0, 1)$ or χ_3^2 .¹¹ In this model, $\alpha_{i0} = \alpha_{i0}(\tau) = \eta_i + F_\epsilon^{-1}(\tau)$ and $\beta_0 = \beta_0(\tau) = 1 + 0.5F_\epsilon^{-1}(\tau)$, where F_ϵ^{-1} is the quantile function of ϵ_{it} . We consider here the cases where $n \in \{50, 100, 500\}$, $T \in \{10, 20, 50, 100\}$ and $\tau \in \{0.25, 0.50, 0.75\}$. We use two different parameter values $\theta \in \{0.5, 1\}$ to control the autocorrelation in the innovation.¹² As in Horowitz (1998), we use the fourth order kernel

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

¹¹The boundedness restriction on the regressors required by condition (A2) is not satisfied by the present specification. However, we believe that condition (A2) is mainly a technical condition and the simulation results should not be affected by the use of this distribution.

¹²In fact, we also examined the cases with a different degree of heteroscedasticity and a different error distribution, but the results are qualitatively similar.

The number of repetitions for each Monte Carlo experiment is 1,000. 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 choose the standard QR estimate as a reasonable starting value. We use `quantreg` package (Koenker, 2009) to compute the standard QR estimates. The estimators under consideration are the standard QR estimator $\hat{\beta}_{\text{KB}}$ 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. Additionally, we apply the proposed one-step bias correction to the standard QR estimator. We denote this estimator as $\hat{\beta}_{\text{KB}}^1$. It is important to note that there is no theoretical justification for this estimator, but it is interesting to analyze its performance in terms of computation. Actually, the computational time of the FE-SQR estimator is considerably slower than the standard QR estimator. Nonetheless, smoothing is crucial to formally characterize the bias expression for panel QR models, so we adopt the smoothing approach here to tackle this important problem.

The choice of the bandwidth is a common, sometimes difficult problem when the kernel smoothing is used. In this experiment, we use $h_{1n} = c_1 s_1 (nT)^{-1/7}$ to construct the FE-SQR estimate, where c_1 is some constant and s_1 is the sample standard deviation of $\tilde{u}_{it} = y_{it} - \hat{\alpha}_{i,\text{KB}} - x_{it}\hat{\beta}_{\text{KB}}$, 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 $\hat{u}_{it} = y_{it} - \hat{\alpha}_i - x_{it}\hat{\beta}$.¹³ 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.¹⁴ On the other hand, by the remark after Theorem 3.2, the bandwidth used to construct $\hat{\mathbf{b}}$, say h_{2n} , should be such that $h_{2n} \asymp T^{-a}$ with $1/9 \leq a < 1/3$ to make $\hat{\mathbf{b}}$ consistent for \mathbf{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 $\hat{\mathbf{b}}$, 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 $c_2 = 2$.

As discussed in the previous section, 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,

¹³ $(\hat{\alpha}_{1,\text{KB}}, \dots, \hat{\alpha}_{n,\text{KB}}, \hat{\beta}_{\text{KB}})$ 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.

¹⁴We do not further discuss the validity of the data dependent bandwidth.

we use the trimming in our implementation, and set $\kappa_n = 0.01$. For the truncation level, we use $m_n = 1$ as discussed in Remark 3.2.

4.2 Results

Before looking at the Monte Carlo results, it is worthwhile to summarize our objectives of this experiment. The objectives are twofold. The first one is to examine whether the bias correction methods work in finite samples. 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 . The second objective is to examine the effect of the bias correction on the precision 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 present the results for the standard deviation (SD), and the root mean squared error (RMSE) of the estimators. The results for $\hat{\beta}_{\text{KB}}$ and $\hat{\beta}_{\text{KB}}^1$ are only for reference, and we mainly focus on the comparison between $\hat{\beta}$, $\hat{\beta}^1$ and $\hat{\beta}_{1/2}$.¹⁵

Table 1 presents the results for bias, SD, and RMSE of the estimators for the normal innovation with $\theta = 0.5$. The first feature from the table is that all estimators are approximately unbiased for $\tau = 0.5$. However, it can be observed that the one-step bias corrected estimator $\hat{\beta}^1$ is able to reduce the bias in $\hat{\beta}$ for $\tau \in \{0.25, 0.75\}$. The estimator $\hat{\beta}_{\text{KB}}^1$ presents similar behavior to $\hat{\beta}^1$ in terms of bias. The results for the half-panel jackknife estimator $\hat{\beta}_{1/2}$ show that for a fixed T , the bias worsens as n increases, but the estimator is able to reduce the bias for large T . The results for SD and RMSE for the normal distribution are reported in the middle and lower parts of Table 1, respectively. The results regarding SD show a inflation of variance of $\hat{\beta}^1$ relative to $\hat{\beta}$ for small T , which, however, decreases as either n or T increases. For small panels, the RMSE of the one-step bias corrected estimator $\hat{\beta}^1$ is slightly larger than those of $\hat{\beta}$. These two observations 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. However, for large n/T ratio, which is common in many applications, the RMSE of the bias corrected estimator is smaller than those of $\hat{\beta}$, for example, for $n = 500$, $T \in \{10, 20\}$ and $\tau \in \{0.25, 0.75\}$, the RMSE's of $\hat{\beta}^1$ are smaller than those of $\hat{\beta}$. This is due to the fact that the bias term dominates the variance in these cases.¹⁶

Table 2 presents the results for bias, SD, and RMSE for the χ_2^3 innovation with $\theta = 0.5$. In this case, for relatively small T , $\hat{\beta}_{\text{KB}}$ and $\hat{\beta}$ are considerably biased.

¹⁵Recall that it is not known whether a result analogous to Theorem 3.1 holds for $\hat{\beta}_{\text{KB}}$.

¹⁶In additional simulation experiments, not reported here but available upon request, we have examined the cases where $n = 1000$ and $T \in \{10, 20\}$. In such cases, the RMSE of the one-step bias corrected estimator $\hat{\beta}^1$ is even smaller than that of $\hat{\beta}$.

However, the bias one-step bias correction is able to substantially reduce the bias. In particular, the bias of $\hat{\beta}$ is the largest when $\tau = 0.75$ among the quantiles under consideration, and $\hat{\beta}^1$ is able to largely remove the bias in this case. The results also show that, for small T and $\tau = 0.75$, the half-panel jackknife estimator $\hat{\beta}_{1/2}$ does not work well. In fact, for $T = 10$ and $\tau = 0.75$, the bias is similar to that of $\hat{\beta}$, the half-panel jackknife method somehow works for large T . The results for SD and RMSE for the χ_2^3 case are presented in the middle and lower parts of Table 2, respectively. They show that both SD and RMSE of $\hat{\beta}^1$ increase substantially for small panel data. However, the variance inflation decreases as either n or T increases, and also as discussed previously, for large n/T ratio the bias term dominates the RMSE and we can observe reduction in RMSE in such cases, for example, for $\tau = 0.75, n = 500$ and $T \in \{10, 20\}$ the RMSE of $\hat{\beta}^1$ are smaller than those of $\hat{\beta}$.

The results for bias, SD, and RMSE for the $\theta = 1$ case are presented in Tables 3 and 4 for the normal and χ_3^2 innovations respectively. Basically, they are parallel to those for $\theta = 0.5$. 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.

As a final remark, we shall comment that, in our simulation experiments, the analytic bias correction presented in Section 3.3 somehow works for the unsmoothed estimator. This is not totally surprising because the two estimator $\hat{\beta}$ and $\hat{\beta}_{\text{KB}}$ naturally share a common structure. In practice, it might be an option to use the bias formula given in Section 3.3 to the unsmoothed estimator. However, it is again noted that there is no formal validity to do so.

5 Concluding Remarks

In this paper, we have shown that the fixed effects 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 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 eliminate the bias in the limiting distribution. Importantly, our results allow for dependence in the time dimension and cover practical situations. 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.

Many issues remain to be investigated. Substantial differences from smooth non-

linear panel models are the facts that the FE-SQR estimator requires the choice of the bandwidth and its asymptotic bias consists of nonparametric objects. In particular, the estimation of the bias term is more complicated than in smooth nonlinear panel models, and the choice of the bandwidth is a topic to be further investigated in future work. Moreover, in this paper, we have suggested a consistent estimator for the asymptotic covariance matrix. However, as observed in the simulation experiments, the inflation of the empirical variance of the analytically bias corrected estimator might be problematic for practical inference. Therefore, extensive numerical studies, such as those of Buchinsky (1995) for the cross section case, will be helpful to determine a good way of making practical inference in panel quantile regression models, which will be pursued in another place.

A Proofs

For any triangular sequence $\{a_{ni}\}_{i=1}^n$ and any sequence $\epsilon_n \rightarrow 0$, we use the notation $a_{ni} = \bar{o}(\epsilon_n)$ if $\max_{1 \leq i \leq n} |a_{ni}| = o(\epsilon_n)$. We also define $\bar{O}(\cdot)$, $\bar{o}_p(\cdot)$ and $\bar{O}_p(\cdot)$ in a similar fashion. For $x, y \in \mathbb{R}$, we use the notation $x \vee y = \max\{x, y\}$. For $x \in \mathbb{R}$, $[x]$ denotes the greatest integer not exceeding x .

A.1 Proof of Proposition 3.1

Put $\mathbb{M}_{ni}(\alpha, \boldsymbol{\beta}) := T^{-1} \sum_{t=1}^T (y_{it} - \alpha - \mathbf{x}'_{it}\boldsymbol{\beta})\{\tau - G_{h_n}(y_{it} - \alpha - \mathbf{x}'_{it}\boldsymbol{\beta})\}$, $\tilde{\mathbb{M}}_{ni}(\alpha, \boldsymbol{\beta}) := T^{-1} \sum_{t=1}^T (y_{it} - \alpha - \mathbf{x}'_{it}\boldsymbol{\beta})\{\tau - I(y_{it} \leq \alpha + \mathbf{x}'_{it}\boldsymbol{\beta})\}$ and $\Delta_i(\alpha, \boldsymbol{\beta}) := E[\tilde{\mathbb{M}}_{ni}(\alpha, \boldsymbol{\beta}) - \tilde{\mathbb{M}}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)]$. 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 $N_\delta := \{(\alpha, \boldsymbol{\beta}) : |\alpha| + \|\boldsymbol{\beta}\|_1 \leq \delta\}$ and N_δ^c by its complement in \mathbb{R}^{p+1} . We first show that for any $\delta > 0$,

$$\inf_{i \geq 1} \min_{(\alpha, \boldsymbol{\beta}) \in N_\delta^c} \Delta_i(\alpha, \boldsymbol{\beta}) > 0. \quad (\text{A.1})$$

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

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

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 $i \geq 1$ as

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

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

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

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

$$\inf_{i \geq 1} \min_{|\alpha| + \|\beta\|_1 = \delta} \Delta_i(\alpha, \beta) > 0. \quad (\text{A.2})$$

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

Next, we shall show that

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

Observe that

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

By the proof of Horowitz (1998, Lemma 1), the first and second terms are $O(h_n)$ uniformly over $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$ and $1 \leq i \leq n$. It remains to prove that third term is $o_p(1)$ uniformly over $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$ and $1 \leq i \leq n$. This can be shown in a similar way to the proof of (A.2) in Kato, Galvao, and Montes-Rojas (2011) (see also their Remark A.1) with a suitable change according to the fact that the observations are now dependent in the time dimension. Since the most of arguments are similar to the proof of their (A.2), we only state the point to be changed. The only point that should be changed is the place where they apply the Marcinkiewicz-Zygmund inequality to evaluate both the terms of their (A.4). Instead of the Marcinkiewicz-Zygmund inequality, we now apply a Bernstein type inequality for β -mixing processes (see Corollary C.1 below) with the evaluation of the variance term by Lemma C.2. Because of the exponential β -mixing property (condition (A1)), and the uniform boundedness of \mathbf{x}_{it} (condition (A2)), taking $s = 2 \log n$ and $q = \lceil \sqrt{T} \rceil$ in Corollary C.1 and using the fact that $(\log n)/\sqrt{T} \rightarrow 0$, it is now shown that for any fixed $\epsilon > 0$, for large n ,

$$\max_{1 \leq i \leq n} \mathbb{P} \left\{ \sup_{(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}} |\tilde{\mathbb{M}}_{ni}(\alpha, \beta) - \tilde{\mathbb{M}}_{ni}(0, \mathbf{0}) - \Delta_i(\alpha, \beta)| > \epsilon \right\} \leq 2 \left(n^{-2} + \sqrt{T} B a^{\lceil \sqrt{T} \rceil} \right).$$

Because $(\log n)/\sqrt{T} \rightarrow 0$, the right side is $o(n^{-1})$. Thus, by the union bound, we have

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

Therefore, we obtain (A.3).

We shall show $\hat{\beta} \xrightarrow{p} \mathbf{0} = \beta_0$. Pick any $\delta > 0$ and put $\epsilon := \inf_{i \geq 1} \min_{(\alpha, \beta) \in N_\delta^c} \Delta_i(\alpha, \beta) > 0$. Observe that when $\|\hat{\beta}\|_1 > \delta$,

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

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

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

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

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

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

Thus, we obtain the inclusion relation

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

By (A.3), $\mathbb{P}(A_{1n}) \rightarrow 0$. On the other hand, since $\Delta_i(0, \hat{\beta}) \leq 2M\|\hat{\beta}\|$, the consistency of $\hat{\beta}$ implies that $\mathbb{P}(A_{2n}) \rightarrow 0$. Therefore, we complete the proof. \square

A.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 $H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}) := \mathbb{E}[\mathbb{H}_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta})]$ and $H_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \mathbb{E}[\mathbb{H}_n^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta})]$. 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 A.1-A.3 below are used to derive a convenient expansion of $\hat{\boldsymbol{\beta}}$.

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

$$\begin{aligned} \hat{\alpha}_i - \alpha_{i0} &= s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - \gamma'_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2^{-1} s_i f_i^{(1)}(0)(\hat{\alpha}_i - \alpha_{i0})^2 \\ &\quad + s_i [\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\}] \\ &\quad + \bar{O}_p(d_n), \end{aligned} \tag{A.4}$$

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 &= \Gamma_n^{-1} \{-n^{-1} \sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) \boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} \\ &\quad - \Gamma_n^{-1} n^{-1} \sum_{i=1}^n \boldsymbol{\gamma}_i [\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\} - \{H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\alpha_{0i}, \boldsymbol{\beta}_0)\}] \\ &\quad + \Gamma_n^{-1} [\{\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} - \{H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\}] \\ &\quad + \Gamma_n^{-1} \{(2n)^{-1} \sum_{i=1}^n \boldsymbol{\nu}_i (\hat{\alpha}_i - \alpha_{i0})^2\} + O_p(d_n). \end{aligned} \tag{A.5}$$

Proof of Lemma A.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

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

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

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

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

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

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

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

Combining (A.8) and (A.6) leads to (A.5). \square

Put

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

Lemma A.2. *We have*

$$\begin{aligned} I_{ni}^{(1)} &= -\{(\hat{f}_i - \mathbb{E}[\hat{f}_i]) - h_n(\tilde{f}_i^{(1)} - \mathbb{E}[\tilde{f}_i^{(1)}])\}(\hat{\alpha}_i - \alpha_{i0}) \\ &\quad + \bar{o}_p[\{(\log n)/(Th_n)^{1/2}\} \delta_{2n} + \{(\log n)/(Th_n^3)^{1/2}\} \delta_{1n}^2], \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} I_n^{(2)} &= -n^{-1} \sum_{i=1}^n \{(\hat{\boldsymbol{g}}_i - \mathbb{E}[\hat{\boldsymbol{g}}_i]) - h_n(\tilde{\boldsymbol{g}}_i^{(1)} - \mathbb{E}[\tilde{\boldsymbol{g}}_i^{(1)}])\}(\hat{\alpha}_i - \alpha_{i0}) \\ &\quad + o_p[\{(\log n)/(Th_n)^{1/2}\} \delta_{2n} + \{(\log n)/(Th_n^3)^{1/2}\} \delta_{1n}^2], \end{aligned} \quad (\text{A.10})$$

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

Proof of Lemma A.2. We only prove (A.9) since the proof of (A.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 B.2, $\partial_{\boldsymbol{\beta}} \mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \tilde{\boldsymbol{\beta}}_i) - \partial_{\boldsymbol{\beta}} H_{ni}^{(1)}(\hat{\alpha}_i, \tilde{\boldsymbol{\beta}}_i) = \bar{o}_p\{(\log n)/(Th_n)^{1/2}\}$, implying that $J_{i1} = \bar{o}_p[\{(\log n)/(Th)^{1/2}\} \delta_{2n}]$. Similarly, use Taylor's theorem to obtain

$$\begin{aligned} J_{i2} &= -\{(\hat{f}_i - \mathbb{E}[\hat{f}_i]) - h_n(\tilde{f}_i^{(1)} - \mathbb{E}[\tilde{f}_i^{(1)}])\}(\hat{\alpha}_i - \alpha_{i0}) \\ &\quad + \{\partial_{\alpha_i}^2 \mathbb{H}_{ni}^{(1)}(\tilde{\alpha}_i, \boldsymbol{\beta}_0) - \partial_{\alpha_i}^2 H_{ni}^{(1)}(\tilde{\alpha}_i, \boldsymbol{\beta}_0)\}(\hat{\alpha}_i - \alpha_{i0})^2, \end{aligned} \quad (\text{A.11})$$

where for each i , $\tilde{\alpha}_i$ is on the line segment between $\hat{\alpha}_i$ and α_{i0} . By Lemma B.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 (A.11) is $\bar{o}_p[\{(\log n)/(Th_n^3)^{1/2}\} \delta_{1n}^2]$. Therefore, we complete the proof. \square

Lemma A.3. $\max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}| = O_p\{(T/\log n)^{-1/2}\}$.

Proof of Lemma A.3. By Lemmas A.2 and B.2, both $n^{-1} \sum_{i=1}^n \gamma_i I_{ni}^{(1)}$ and $I_n^{(2)}$ are of order $o_p\{\max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}| + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|\}$. Thus, by (A.5),

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 &= \Gamma_n^{-1} \{-n^{-1} \sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) \gamma_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} \\ &\quad + o_p\{\max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}| + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|\}. \end{aligned}$$

By the exponential β -mixing property (condition (A1)), the uniform boundedness of \boldsymbol{x}_{it} , and the discussion following Corollary C.2 below it is shown that the variance of the term $n^{-1} \sum_{i=1}^n \{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) \gamma_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\}$ is $O\{(nT)^{-1}\}$. Thus, combining the fact that the expectation of that term is $O(h_n^r)$ by Lemma B.1, we have

$$\begin{aligned} \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| &= O_p\{(nT)^{-1/2} + h_n^r\} + o_p\{\max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}|\} \\ &= o_p\{T^{-1/2} + \max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}|\}. \end{aligned}$$

Using this bound, together with Lemma A.2, we obtain by (A.4)

$$\hat{\alpha}_i - \alpha_{i0} = s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) + \bar{o}_p\{T^{-1/2} + \max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}|\}.$$

We now bound the term $\max_{1 \leq i \leq n} |\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)|$. Observe that

$$\begin{aligned} \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) &= \{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - H_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)\} + H_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0) \\ &= \{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - H_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)\} + \bar{O}(h_n^r). \end{aligned}$$

We wish to show that

$$\max_{1 \leq i \leq n} |\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - H_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)| = O_p\{(T/\log n)^{-1/2}\}. \quad (\text{A.12})$$

Because of the fact that each term in the sum $\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)$ is uniformly bounded by some constant independent of i and n , applying Corollary C.1 to that sum with $s = 2 \log n$ and $q = \lceil (T/\log n)^{1/2} \rceil$, and using Lemma C.2 to evaluate the variance term, we have

$$\begin{aligned} &\mathbb{P}\left\{|\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - H_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)| \geq \text{const.} \times (T/\log n)^{-1/2}\right\} \\ &\leq 2 \left\{n^{-2} + (T \log n)^{1/2} B a^{\lceil (T/\log n)^{1/2} \rceil}\right\}, \end{aligned}$$

where the constant is independent of i and n . Since the right side is $\bar{o}(n^{-1})$, we obtain (A.12) by the union bound. Therefore, we obtain the desired result. \square

By Lemmas A.1-A.3,

$$\begin{aligned}
\hat{\alpha}_i - \alpha_{i0} &= s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - \boldsymbol{\gamma}'_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2^{-1} s_i f_i^{(1)}(0)(\hat{\alpha}_i - \alpha_{i0})^2 \\
&\quad - s_i \{(\hat{f}_i - \mathbb{E}[\hat{f}_i]) - h_n(\tilde{f}_i^{(1)} - \mathbb{E}[\tilde{f}_i^{(1)}])\}(\hat{\alpha}_i - \alpha_{i0}) + \bar{o}_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) \\
&= s_i \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) + \bar{O}_p\{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + (\log n)^{3/2}/(Th_n^{1/2})\}, \tag{A.13}
\end{aligned}$$

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

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

Plugging (A.13) into (A.9) and (A.10), and using Lemma B.2, we obtain

$$n^{-1} \sum_{i=1}^n \boldsymbol{\gamma}_i I_{in}^{(1)} = -B_1 + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|), \quad I_n^{(2)} = -B_2 + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|). \tag{A.14}$$

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

$$n^{-1} \sum_{i=1}^n \boldsymbol{\nu}_i (\hat{\alpha}_i - \alpha_{i0})^2 = B_3 + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|). \tag{A.15}$$

Plugging (A.14)-(A.15) into (A.5), we obtain

$$\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 &= \Gamma_n^{-1} \{-n^{-1} \sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) \boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} \\
&\quad + \Gamma_n^{-1} (B_1 - B_2 + B_3) + o_p(T^{-1} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|). \tag{A.16}
\end{aligned}$$

The first term on the right side of (A.16) is expected to follow a central limit theorem. We will establish this fact in Lemma A.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 A.5 and A.6.

Lemma A.4. *Recall the definition of V given in condition (A8). We have*

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

Proof of Lemma A.4. For any fixed $\mathbf{c} \in \mathbb{R}^p$, put $z_{it} := \{\tau - G_{h_n}(u_{it}) + h_n \tilde{K}_{h_n}(u_{it})\} \mathbf{c}'(\mathbf{x}_{it} - \boldsymbol{\gamma}_i)$. Note that z_{it} depends on n . By the Cramér-Wold device, it suffices to show that $(nT)^{-1/2} \sum_{i=1}^n \sum_{t=1}^T z_{it} \xrightarrow{d} N(0, \mathbf{c}' V \mathbf{c})$. Observe that by Lemma B.1,

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T z_{it} = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (z_{it} - \mathbb{E}[z_{i1}]) + O\{(nT)^{1/2} h_n^r\},$$

and the second term is $o(1)$ because of the fact that $h_n^r = o(T^{-1}) = o\{(nT)^{-1/2}\}$.

Put $\tilde{z}_{it} := z_{it} - \mathbb{E}[z_{i1}]$. We first evaluate the variance of the term $(nT)^{-1/2} \sum_{i=1}^n \sum_{t=1}^T \tilde{z}_{it}$. By a standard calculation,

$$\mathbb{E}[\tilde{z}_{i1}^2] = \mathbb{E}[\{\{\tau - G_{h_n}(u_{i1})\}\mathbf{c}'(\mathbf{x}_{i1} - \boldsymbol{\gamma}_i)\}^2] + \bar{O}(h_n).$$

and uniformly over both \mathbf{x} and i ,

$$\begin{aligned} & \mathbb{E}[\{\tau - G_{h_n}(u_{i1})\}^2 | \mathbf{x}_{i1} = \mathbf{x}] \\ &= \tau^2 - 2\tau \mathbb{E}[G_{h_n}(u_{i1}) | \mathbf{x}_{i1} = \mathbf{x}] + \mathbb{E}[G_{h_n}(u_{i1})^2 | \mathbf{x}_{i1} = \mathbf{x}] \\ &= -\tau^2 + O(h_n^r) + \int_{-\infty}^{\infty} \int_{u/h_n}^{\infty} \int_{u/h_n}^{\infty} K(v_1)K(v_2)dv_1dv_2f_i(u|\mathbf{x})du \\ &= -\tau^2 + O(h_n^r) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i(h_n \min\{v_1, v_2\} | \mathbf{x})K(v_1)K(v_2)dv_1dv_2 \\ &= \tau(1 - \tau) + O(h_n). \end{aligned} \tag{A.17}$$

On the other hand, uniformly over (i, j) ($j \geq 1$),

$$\text{Cov}(z_{i1}, z_{i,1+j}) = \mathbb{E}[\{\tau - G_{h_n}(u_{i1})\}\{\tau - G_{h_n}(u_{i,1+j})\}\mathbf{c}'(\mathbf{x}_{i1} - \boldsymbol{\gamma}_i)\mathbf{c}'(\mathbf{x}_{i,1+j} - \boldsymbol{\gamma}_i)] + O(h_n^r),$$

and uniformly over both $(\mathbf{x}_1, \mathbf{x}_{1+j})$ and (i, j) ($j \geq 1$),

$$\begin{aligned} & \mathbb{E}[\{\tau - G_{h_n}(u_{i1})\}\{\tau - G_{h_n}(u_{i,1+j})\} | \mathbf{x}_{i1} = \mathbf{x}_1, \mathbf{x}_{i,1+j} = \mathbf{x}_{1+j}] \\ &= -\tau^2 + O(h_n^r) + \mathbb{E}[G_{h_n}(u_{i1})G_{h_n}(u_{i,1+j}) | \mathbf{x}_{i1} = \mathbf{x}_1, \mathbf{x}_{i,1+j} = \mathbf{x}_{1+j}] \\ &= -\tau^2 + O(h_n^r) \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{u_1/h_n}^{\infty} \int_{u_{1+j}/h_n}^{\infty} K(v_1)K(v_2)dv_1dv_2f_{i,j}(u_1, u_{1+j} | \mathbf{x}_1, \mathbf{x}_{1+j})du_1du_{1+j} \\ &= -\tau^2 + O(h_n^r) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{i,j}(h_nv_1, h_nv_2 | \mathbf{x}_1, \mathbf{x}_{1+j})K(v_1)K(v_2)dv_1dv_2 \\ &= -\tau^2 + F_{i,j}(0, 0 | \mathbf{x}_1, \mathbf{x}_{1+j}) + O(h_n^r) \\ &= \mathbb{E}[\{\tau - I(u_{i1} \leq 0)\}\{\tau - I(u_{i,1+j} \leq 0)\} | \mathbf{x}_{i1} = \mathbf{x}_1, \mathbf{x}_{i,1+j} = \mathbf{x}_{1+j}] + O(h_n^r), \end{aligned} \tag{A.18}$$

where the fourth equality is due to the fact that $K(\cdot)$ is an r -th order kernel. Put $\tilde{z}_{it} := \{\tau - I(u_{it} \leq 0)\}\mathbf{c}'(\mathbf{x}_{it} - \boldsymbol{\gamma}_i)$. Then, by (A.17) and (A.18), $\text{Var}(z_{i1}) = \text{Var}(\tilde{z}_{i1}) + \bar{O}(h_n)$ and $\text{Cov}(z_{i1}, z_{i,1+j}) = \text{Cov}(\tilde{z}_{i1}, \tilde{z}_{i,1+j}) + O(h_n^r)$ uniformly over (i, j) ($j \geq 1$). Thus,

$$\begin{aligned} \text{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_{it}\right) &= \sum_{|j| \leq T-1} \left(1 - \frac{|j|}{T}\right) \text{Cov}(z_{i1}, z_{i,1+j}) \\ &= \sum_{|j| \leq T-1} \left(1 - \frac{|j|}{T}\right) \text{Cov}(\tilde{z}_{i1}, \tilde{z}_{i,1+j}) + \bar{O}(h_n + Th_n^r) \\ &= \text{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{z}_{it}\right) + \bar{O}(h_n + Th_n^r), \end{aligned}$$

and $h_n + Th_n^r \rightarrow 0$. Therefore, we have

$$\begin{aligned} \text{Var} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T z_{it} \right) &= \frac{1}{n} \sum_{i=1}^n \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_{it} \right) = \frac{1}{n} \sum_{i=1}^n \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{z}_{it} \right) + o(1) \\ &= \mathbf{c}' \left(\frac{1}{n} \sum_{i=1}^n V_{ni} \right) \mathbf{c} + o(1) \rightarrow \mathbf{c}' V \mathbf{c}. \end{aligned}$$

We wish to show a central limit theorem for the term $(nT)^{-1/2} \sum_{i=1}^n \sum_{t=1}^T \tilde{z}_{it}$. Put $\bar{z}_{i+} := T^{-1/2} \sum_{t=1}^T \tilde{z}_{it}$. Observe that $\bar{z}_{1+}, \dots, \bar{z}_{n+}$ are independent. Viewing that

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \tilde{z}_{it} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{z}_{i+},$$

we check the Lyapunov condition for the right sum. By the previous result, it suffices to show that $\sum_{i=1}^n \mathbb{E}[|\bar{z}_{i+}|^3] = o(n^{3/2})$. By conditions (A2) and (A6), \tilde{z}_{it} is uniformly bounded. By the exponential β -mixing property (condition (A1)) and Theorem 3 of Yoshihara (1978), it is shown that $\mathbb{E}[|\bar{z}_{i+}|^3] = \bar{O}(1)$, which implies that $\sum_{i=1}^n \mathbb{E}[|\bar{z}_{i+}|^3] = O(n) = o(n^{3/2})$.¹⁷ Therefore, by the Lyapunov central limit theorem, we have $n^{-1/2} \sum_{i=1}^n \bar{z}_{i+} \xrightarrow{d} N(0, \mathbf{c}' V \mathbf{c})$. \square

Lemma A.5. Recall the definition of $\omega_{ni}^{(3)}$ given in condition (A9). B_3 is expanded as

$$B_3 = \frac{1}{T} \cdot \left(\frac{1}{2n} \sum_{i=1}^n s_i^2 \omega_{ni}^{(3)} \boldsymbol{\nu}_i \right) + o_p(T^{-1}).$$

Proof of Lemma A.5. Put $z_{it} := \tau - G_{h_n}(u_{it}) + h_n \tilde{K}_{h_n}(u_{it})$ and $\bar{z}_{it} := z_{it} - \mathbb{E}[z_{i1}]$. Then, by Lemma B.1 and the proof of the previous lemma, we have

$$\begin{aligned} \mathbb{E}[\{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\}^2] &= \mathbb{E}[\{\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - \mathbb{E}[\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)]\}^2] + (\mathbb{E}[\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)])^2 \\ &= \frac{1}{T} \left\{ \sum_{|j| \leq T-1} \left(1 - \frac{|j|}{T} \right) \text{Cov}(z_{i1}, z_{i,1+j}) \right\} + \bar{O}(h_n^{2r}) \\ &= \frac{1}{T} \left[\sum_{|j| \leq T-1} \left(1 - \frac{|j|}{T} \right) \text{Cov}\{I(u_{i1} \leq 0), I(u_{i,1+j} \leq 0)\} \right] + \bar{o}(T^{-1}) \\ &= T^{-1} \omega_{ni}^{(1)} + \bar{o}(T^{-1}). \end{aligned}$$

Thus, $\mathbb{E}[B_3] = (n^{-1} \sum_{i=1}^n s_i^2 \omega_{ni}^{(3)} \boldsymbol{\nu}_i) / (2T) + o(T^{-1})$.

¹⁷Theorem 3 of Yoshihara (1978) is stated in terms of α -mixing processes. However, since each α -mixing coefficient is bounded by the corresponding β -mixing coefficient, Theorem 3 of Yoshihara (1978) holds with α -mixing coefficients replaced by β -mixing coefficients.

We next evaluate the variance of any linear combination of B_3 . For any fixed $\mathbf{c} \in \mathbb{R}^p$, put $a_i := s_i^2 \mathbf{c}' \boldsymbol{\nu}_i$. Then,

$$\begin{aligned}
& \text{Var}\{n^{-1} \sum_{i=1}^n a_i (T^{-1} \sum_{t=1}^T z_{it})^2\} \\
&= n^{-2} \sum_{i=1}^n a_i^2 \text{Var}\{(T^{-1} \sum_{t=1}^T z_{it})^2\} \\
&\leq n^{-2} \sum_{i=1}^n a_i^2 \mathbb{E}[(T^{-1} \sum_{t=1}^T z_{it})^4] \\
&\leq 8n^{-2} \sum_{i=1}^n a_i^2 \mathbb{E}[(T^{-1} \sum_{t=1}^T \bar{z}_{it})^4] + O(n^{-1} h_n^{4r}). \tag{A.19}
\end{aligned}$$

Since $|\bar{z}_{it}|$ is uniformly bounded by some constant depending only on $K(\cdot)$, by the exponential β -mixing property (condition (A1)) and Theorem 3 of Yoshihara (1978), it is shown that the first term on (A.19) is $O(n^{-1} T^{-2})$. Therefore, we obtain the desired result. \square

Lemma A.6. Recall the definitions of $\omega_{ni}^{(2)}$ and $\omega_{ni}^{(3)}$ given in condition (A9). B_1 and B_2 are expanded as

$$\begin{aligned}
B_1 &= \frac{1}{T} \cdot \left(\frac{2\tau - 1}{2n} \sum_{i=1}^n \gamma_i + \frac{1}{n} \sum_{i=1}^n s_i \omega_{ni}^{(1)} \gamma_i \right) + o_p(T^{-1}), \\
B_2 &= \frac{1}{T} \cdot \left(\frac{2\tau - 1}{2n} \sum_{i=1}^n \gamma_i + \frac{1}{n} \sum_{i=1}^n s_i \omega_{ni}^{(2)} \right) + o_p(T^{-1}).
\end{aligned}$$

Proof of Lemma A.6. We only prove the expansion for B_1 . The expansion for B_2 can be shown in a similar way. Put $z_{it} := \tau - G_{h_n}(u_{it}) + h_n \tilde{K}_{h_n}(u_{it})$ and $\bar{z}_{it} := z_{it} - \mathbb{E}[z_{it}]$. Recall that $\mathbb{E}[z_{i1}] = \bar{O}(h_n^r)$ by Lemma B.1. Observe that

$$\begin{aligned}
\mathbb{E}[(\hat{f}_i - \mathbb{E}[\hat{f}_i])(T^{-1} \sum_{t=1}^T z_{it})] &= \mathbb{E}[\hat{f}_i(T^{-1} \sum_{t=1}^T z_{it})] + \bar{O}(h_n^r) \\
&= T^{-2} \sum_{s=1}^T \sum_{t=1}^T \mathbb{E}[K_{h_n}(u_{is}) z_{it}] + O(h_n^r),
\end{aligned}$$

We evaluate the term $\mathbb{E}[K_{h_n}(u_{is}) z_{it}]$. First,

$$\begin{aligned}
\mathbb{E}[K_{h_n}(u_{i1}) z_{i1}] &= \int_{-\infty}^{\infty} K(u) \{\tau - G(u)\} f_i(uh_n) du + \int_{-\infty}^{\infty} u K(u)^2 f_i(uh_n) du \\
&= f_i(0) \left\{ \tau - \int_{-\infty}^{\infty} K(u) G(u) du + \int_{-\infty}^{\infty} u K(u)^2 du \right\} + \bar{O}(h_n).
\end{aligned}$$

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

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

Thus, we have $E[K_{h_n}(u_{i1})z_{i1}] = (\tau - 1/2)f_i(0) + \bar{O}(h_n)$. Second, for $|j| \geq 1$,

$$\begin{aligned}
E[K_{h_n}(u_{i1})z_{i,1+j}] &= \tau \int_{-\infty}^{\infty} K(u)f_i(uh_n)du \\
&\quad - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(u_1)G_{h_n}(u_{1+j})f_{i,j}(u_1h_n, u_{1+j})du_1du_{1+j} \\
&\quad + h_n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(u_1)\tilde{K}(u_{1+j})f_{i,j}(u_1h_n, u_{1+j}h_n)du_1du_{1+j} \\
&= \tau f_i(0) + O(h_n^r) \\
&\quad - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{h_nv} f_{i,j}(u_1h_n, u_{1+j})du_{1+j} \right\} K(u_1)K(v)du_1dv \\
&= \tau f_i(0) - \int_{-\infty}^0 f_{i,j}(0, u)du + O(h_n^r),
\end{aligned}$$

where the $O(h_n^r)$ term is independent of i and j . Therefore, we have

$$E[(\hat{f}_i - E[\hat{f}_i])(T^{-1} \sum_{t=1}^T z_{it})] = T^{-1}(\tau - 1/2)f_i(0) + T^{-1}\omega_{ni}^{(2)} + \bar{o}(T^{-1}). \quad (\text{A.20})$$

Similarly, observe that

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

We evaluate the term $h_n^{-2}E[\tilde{K}^{(1)}(u_{is}/h_n)z_{it}]$. First,

$$\begin{aligned}
h_n^{-2}E[\tilde{K}^{(1)}(u_{i1}/h_n)z_{i1}] &= h_n^{-1} \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u)\{\tau - G(u)\}f_i(uh_n)du \\
&\quad + h_n^{-1} \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u)\tilde{K}(u)f_i(uh_n)du \\
&= -h_n^{-1} \int_{-\infty}^{\infty} \tilde{K}(u)K(u)f_i(uh_n)du + \bar{O}(1) \\
&= -h_n^{-1}f_i(0) \int_{-\infty}^{\infty} uK(u)^2du + \bar{O}(1) = \bar{O}(1),
\end{aligned}$$

where the second equality is due to the fact that

$$\begin{aligned}
\int_{-\infty}^{\infty} \tilde{K}^{(1)}(u)\tilde{K}(u)f_i(uh_n)du &= - \int_{-\infty}^{\infty} \tilde{K}(u)\tilde{K}^{(1)}(u)f_i(uh_n)du - h_n \int_{-\infty}^{\infty} \tilde{K}^2(u)f_i^{(1)}(uh_n)du \\
&= - \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u)\tilde{K}(u)f_i(uh_n)du + \bar{O}(h_n).
\end{aligned}$$

Second, for $|j| \geq 1$, we have

$$\begin{aligned}
h_n^{-2} \mathbf{E}[\tilde{K}^{(1)}(u_{i1}/h_n) z_{i,1+j}] &= h_n^{-1} \tau \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u) f_i(u h_n) du \\
&\quad - h_n^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u_1) G_{h_n}(u_{1+j}) f_{i,j}(u_1 h_n, u_{1+j}) du_1 du_{1+j} \\
&\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{K}^{(1)}(u_1) \tilde{K}(u_{1+j}) f_{i,j}(u_1 h_n, u_{1+j} h_n) du_1 du_{1+j} \\
&= -\tau \int_{-\infty}^{\infty} \tilde{K}(u) f_i^{(1)}(u h_n) du \\
&\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{K}(u_1) K(v) \left\{ \int_{-\infty}^{v h_n} f_{i,j}^{(1,0)}(u_1 h_n, u_{1+j}) du_{1+j} \right\} du_1 dv \\
&\quad - h_n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{K}(u_1) \tilde{K}(u_{1+j}) f_{i,j}^{(1,0)}(u_1 h_n, u_{1+j} h_n) du_1 du_{1+j} \\
&= O(h_n^{r-1}),
\end{aligned}$$

where the $O(h_n^{r-1})$ term is independent of i and j . Therefore, we have

$$h_n \mathbf{E}[(\tilde{f}_i^{(1)} - \mathbf{E}[\tilde{f}_i^{(1)}])(T^{-1} \sum_{t=1}^T z_{it})] = \bar{O}(T^{-1} h_n + h_n^r) = \bar{o}(T^{-1}). \quad (\text{A.21})$$

Combining (A.20) and (A.21), we obtain

$$\mathbf{E}[B_1] = \frac{1}{T} \cdot \left(\frac{2\tau - 1}{2n} \sum_{i=1}^n \gamma_i + \frac{1}{n} \sum_{i=1}^n s_i \omega_{ni}^{(1)} \gamma_i \right) + o(T^{-1}). \quad (\text{A.22})$$

It remains to evaluate the variance of any linear combination of B_1 . For any fixed $\mathbf{c} \in \mathbb{R}^p$, put $a_i := s_i \mathbf{c}' \gamma_i$. Then,

$$\begin{aligned}
&\text{Var}\{n^{-1} \sum_{i=1}^n a_i (\hat{f}_i - \mathbf{E}[\hat{f}_i])(T^{-1} \sum_{t=1}^T z_{it})\} \\
&= n^{-2} \sum_{i=1}^n a_i^2 \text{Var}\{(\hat{f}_i - \mathbf{E}[\hat{f}_i])(T^{-1} \sum_{t=1}^T z_{it})\} \\
&\leq n^{-2} \sum_{i=1}^n a_i^2 \mathbf{E}[(\hat{f}_i - \mathbf{E}[\hat{f}_i])^2 (T^{-1} \sum_{t=1}^T z_{it})^2] \\
&\leq 2n^{-2} \sum_{i=1}^n a_i^2 \mathbf{E}[(\hat{f}_i - \mathbf{E}[\hat{f}_i])^2 \{T^{-1} \sum_{t=1}^T (z_{it} - \mathbf{E}[z_{it}])\}^2] \\
&\quad + 2n^{-2} \sum_{i=1}^n a_i^2 (\mathbf{E}[z_{i1}])^2 \mathbf{E}[(\hat{f}_i - \mathbf{E}[\hat{f}_i])^2].
\end{aligned} \quad (\text{A.23})$$

By the proof of Lemma B.2 below, $\mathbf{E}[(\hat{f}_i - \mathbf{E}[\hat{f}_i])^2] = \bar{O}\{(Th_n)^{-1}\}$. Thus, by Lemma B.1, the second term on the right side of (A.23) is $O(n^{-1} T^{-1} h_n^{2r-1})$. Put $w_{it} := K(u_{it}/h_n)$, $\bar{w}_{it} := w_{it} - \mathbf{E}[w_{i1}]$ and $\bar{z}_{it} := z_{it} - \mathbf{E}[z_{i1}]$. Observe that

$$\mathbf{E}[(\hat{f}_i - \mathbf{E}[\hat{f}_i])^2 \{T^{-1} \sum_{t=1}^T (z_{it} - \mathbf{E}[z_{it}])\}^2] = T^{-4} h_n^{-2} \sum_{s=1}^T \sum_{t=1}^T \sum_{u=1}^T \sum_{v=1}^T \mathbf{E}[\bar{w}_{is} \bar{w}_{it} \bar{z}_{iu} \bar{z}_{iv}].$$

It is standard to show that $E[\bar{w}_{i1}^2 \bar{z}_{i1}^2] = \bar{O}(h_n)$. We wish to show that

$$\begin{aligned} \sum_{s \neq t} E[\bar{w}_{is} \bar{w}_{it} \bar{z}_{it}^2] &= \bar{O}(T^2 h_n), \quad \sum_{s \neq t} E[\bar{z}_{is} \bar{w}_{it}^2 \bar{z}_{it}] = \bar{O}(T^2 h_n), \\ \sum_{s \neq t} E[\bar{w}_{is}^2 \bar{z}_{it}^2] &= \bar{O}(T^2 h_n), \quad \sum_{s \neq t} E[\bar{w}_{is} \bar{z}_{is} \bar{w}_{it} \bar{z}_{it}] = \bar{O}(T^2 h_n), \\ \sum_{s,t,u:\text{distinct}} E[\bar{w}_{is}^2 \bar{z}_{it} \bar{z}_{iu}] &= \bar{O}(T^2 h_n), \quad \sum_{s,t,u:\text{distinct}} E[\bar{w}_{is} \bar{w}_{it} \bar{z}_{it} \bar{z}_{iu}] = \bar{O}(T^2 h_n), \\ \sum_{s,t,u:\text{distinct}} E[\bar{w}_{is} \bar{w}_{it} \bar{z}_{iu}^2] &= \bar{O}(T^2 h_n), \quad \sum_{s,t,u,v:\text{distinct}} E[\bar{w}_{is} \bar{w}_{it} \bar{z}_{iu} \bar{z}_{iv}] = \bar{O}(T^2 h_n). \end{aligned}$$

The first four equations are relatively standard to derive. Thus, we concentrate on deriving the last four equations. The evaluation below is similar to the proof of Yoshihara (1978, Theorem 1), but the situation is slightly different from his. For the sake of completeness, we provide a proof of the last four equations. In what follows, C denotes some constant independent of s, t, u, v and i, n . Its value may change from line to line.

Evaluation of $\sum_{s,t,u:\text{distinct}} E[\bar{w}_{is}^2 \bar{z}_{it} \bar{z}_{iu}]$: We first show that for any $\delta > 0$,

$$\sum_{s < t < u} E[\bar{w}_{is}^2 \bar{z}_{it} \bar{z}_{iu}] = \bar{O}\{(Th_n^{1/(1+\delta)}) \vee (T^2 h_n)\}. \quad (\text{A.24})$$

Consider the case that $s < t < u$ and $u - t > t - s$. In that case, $E[|\bar{w}_{is}^2 \bar{z}_{it}|^{1+\delta}] \leq Ch_n$, $E[|\bar{z}_{iu}|^{1+\delta}] \leq C$ and $E[|\bar{w}_{is}^2 \bar{z}_{it} \bar{z}_{iu}|^{1+\delta}] \leq Ch_n$. Thus, since $E[\bar{z}_{iu}] = 0$, by Lemma 1 of Yoshihara (1976),

$$\begin{aligned} \sum_{\substack{s < t < u \\ u-t > t-s}} |E[\bar{w}_{is}^2 \bar{z}_{it} \bar{z}_{iu}]| &\leq Ch_n^{1/(1+\delta)} \sum_{\substack{s < t < u \\ u-t > t-s}} \beta_i(u-t)^{\delta/(1+\delta)} \\ &\leq Ch_n^{1/(1+\delta)} \sum_{s=1}^{T-2} \sum_{j=2}^{T-s-1} \sum_{k=1}^{j-1} \beta_i(j)^{\delta/(1+\delta)} \quad (j = u-t; k = t-s) \\ &\leq CT h_n^{1/(1+\delta)} \sum_{j=1}^{\infty} j \beta_i(j)^{\delta/(1+\delta)} \\ &\leq CT h_n^{1/(1+\delta)}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \sum_{\substack{s < t < u \\ t-s \geq u-t}} |E[\bar{w}_{is}^2 \bar{z}_{it} \bar{z}_{iu}]| &\leq E[\bar{w}_{i1}^2] \sum_{\substack{s < t < u \\ t-s \geq u-t}} |E[\bar{z}_{it} \bar{z}_{iu}]| + \sum_{\substack{s < t < u \\ t-s \geq u-t}} |E[(\bar{w}_{is}^2 - E[\bar{w}_{i1}^2]) \bar{z}_{it} \bar{z}_{iu}]| \\ &\leq Ch_n \sum_{\substack{s < t < u \\ t-s \geq u-t}} \beta_i(u-t)^{\delta/(1+\delta)} + CT h_n^{1/(1+\delta)} \\ &\leq Ch_n \sum_{s=1}^{T-2} \sum_{j=1}^{T-s-1} \sum_{k=1}^j \beta_i(k)^{\delta/(1+\delta)} + CT h_n^{1/(1+\delta)} \\ &\leq CT^2 h_n + CT h_n^{1/(1+\delta)}. \end{aligned}$$

Therefore, we obtain (A.24). Since $\delta > 0$ is arbitrary, taking δ sufficiently small, we have $Th_n^{1/(1+\delta)} = O(T^2h_n)$. By repeating the same argument, we have

$$\sum_{s,t,u:\text{distinct}} \mathbb{E}[\bar{w}_{is}^2 \bar{z}_{it} \bar{z}_{iu}] = \bar{O}(T^2h_n).$$

Evaluation of $\sum_{s,t,u:\text{distinct}} \mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{it}\bar{z}_{iu}]$ and $\sum_{s,t,u:\text{distinct}} \mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}^2]$: By using the same argument as in the previous case, we have

$$\sum_{s,t,u:\text{distinct}} \mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{it}\bar{z}_{iu}] = \bar{O}(T^2h_n), \quad \sum_{s,t,u:\text{distinct}} \mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}^2] = \bar{O}(T^2h_n).$$

Evaluation of $\sum_{s,t,u,v:\text{distinct}} \mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}]$: Consider the case that $s < t < u < v$ and $t - s \geq \max\{u - t, v - u\}$. In that case, for any $\delta > 0$, $\mathbb{E}[|w_{is}|^{1+\delta}] \leq Ch_n$, $\mathbb{E}[|\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}|^{1+\delta}] \leq Ch_n$ and $\mathbb{E}[|\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}|^{1+\delta}] \leq Ch_n^2$. Thus, by Lemma 1 of Yoshihara (1976),

$$\begin{aligned} \sum_{\substack{s < t < u < v \\ t-s \geq \max\{u-t, v-u\}}} |\mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}]| &\leq Ch_n^{2/(1+\delta)} \sum_{\substack{s < t < u < v \\ t-s \geq \max\{u-t, v-u\}}} \beta_i(t-s)^{\delta/(1+\delta)} \\ &\leq Ch_n^{2/(1+\delta)} \sum_{s=1}^{T-3} \sum_{j=1}^{T-s-2} \sum_{k=1}^j \sum_{l=1}^j \beta_i(j)^{\delta/(1+\delta)} \\ &\leq CTh_n^{2/(1+\delta)} \sum_{j=1}^{\infty} j^2 \beta_i(j)^{\delta/(1+\delta)} \\ &\leq CTh_n^{2/(1+\delta)}. \end{aligned}$$

Similarly, we have

$$\sum_{\substack{s < t < u < v \\ v-u \geq \max\{t-s, u-t\}}} |\mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}]| \leq CTh_n^{2/(1+\delta)}.$$

Consider now the case that $s < t < u < v$ and $u - t \geq \max\{t - s, v - u\}$. In that case, by Lemma 1 of Yoshihara (1976), for any $\delta > 0$,

$$\begin{aligned} |\mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}]| &\leq |\text{Cov}(\bar{w}_{is}\bar{w}_{it}, \bar{z}_{iu}\bar{z}_{iv})| + |\mathbb{E}[\bar{w}_{is}\bar{w}_{it}]\mathbb{E}[\bar{z}_{iu}\bar{z}_{iv}]| \\ &\leq Ch_n^{2/(1+\delta)} \beta_i(u-t)^{\delta/(1+\delta)} + Ch_n^{2/(1+\delta)} \beta_i(t-s)^{\delta/(1+\delta)} \beta_i(v-u)^{\delta/(1+\delta)}. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{\substack{s < t < u < v \\ u-t \geq \max\{t-s, v-u\}}} |\mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}]| &\leq CTh_n^{2/(1+\delta)} + Ch_n^{2/(1+\delta)} \sum_{s=1}^{T-3} \sum_{j=1}^{T-s-2} \sum_{k=1}^j \sum_{l=1}^j \beta_i(k)^{\delta/(1+\delta)} \beta_i(l)^{\delta/(1+\delta)} \\ &\leq CTh_n^{2/(1+\delta)} + CT^2h_n^{2/(1+\delta)}. \end{aligned}$$

Taking $\delta \in (0, 1]$, we have

$$\sum_{s < t < u < v} \mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}] = \bar{O}(T^2 h_n).$$

By repeating the same argument, we have

$$\sum_{s,t,u,v:\text{distinct}} \mathbb{E}[\bar{w}_{is}\bar{w}_{it}\bar{z}_{iu}\bar{z}_{iv}] = \bar{O}(T^2 h_n).$$

We have shown that the first term on the right side of (A.23) is $O(n^{-1}T^{-4}h_n^{-2} \times T^2 h_n) = O(n^{-1}T^{-2}h_n^{-1})$. Therefore, $\text{Var}(\mathbf{c}'B_1) = O(n^{-1}T^{-2}h_n^{-1} + n^{-1}T^{-1}h_n^{2r-1}) = o\{(nT)^{-1}\}$. Combining (A.22), we obtain the desired result. \square

By (A.16) and Lemmas A.4-A.6, $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\{(nT)^{-1/2} + T^{-1}\} = O_p\{(nT)^{-1/2}\}$, and

$$\begin{aligned} \sqrt{nT} \cdot \Gamma_n^{-1}(B_1 - B_2 + B_3) &= \sqrt{\frac{n}{T}} \Gamma_n^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n s_i \left(\omega_{ni}^{(1)} \boldsymbol{\gamma}_i - \omega_{ni}^{(2)} + \frac{s_i \omega_{ni}^{(3)} \boldsymbol{\nu}_i}{2} \right) \right\} + o_p(1) \\ &\xrightarrow{p} \sqrt{\rho} \mathbf{b}, \end{aligned}$$

where \mathbf{b} is given in (3.2). Therefore, we obtain (3.1). \square

A.3 Proof of Theorem 3.2

It suffices to show that $\hat{\mathbf{b}} \xrightarrow{p} \mathbf{b}$. To this end, we first show that

$$\max_{1 \leq i \leq n} |\hat{s}_i - s_i| \xrightarrow{p} 0, \quad \max_{1 \leq i \leq n} \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\| \xrightarrow{p} 0, \quad \hat{\Gamma}_n^{-1} = \Gamma_n^{-1} + o_p(1). \quad (\text{A.25})$$

We use the notation in Appendix B 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 B.2, we have $\hat{f}_i(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) = \mathbb{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})]_{\alpha=\hat{\alpha}_i, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \bar{o}_p\{(\log n)/(Th_n)^{1/2}\}$. Observe that $\mathbb{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})] = \{\mathbb{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})] - \mathbb{E}[\hat{f}_i(0, \mathbf{0})]\} + \mathbb{E}[\hat{f}_i(0, \mathbf{0})]$, $\mathbb{E}[\hat{f}_i(0, \mathbf{0})] = f_i(0) + \bar{O}(h_n^r) = f_i(0) + \bar{o}_p\{(\log n)/(Th_n)^{1/2}\}$ and

$$\begin{aligned} &|\mathbb{E}[\hat{f}_i(\alpha, \boldsymbol{\beta})] - \mathbb{E}[\hat{f}_i(0, \mathbf{0})]| \\ &\leq \mathbb{E} \left[\int_{-\infty}^{\infty} |K(u) \{f_i(uh_n + \alpha + \mathbf{x}'_{i1} \boldsymbol{\beta} | \mathbf{x}_{i1}) - f_i(uh_n | \mathbf{x}_{i1})\}| du \right] \\ &\leq C_f (|\alpha| + M \|\boldsymbol{\beta}\|) \int_{-\infty}^{\infty} |K(u)| du. \end{aligned} \quad (\text{A.26})$$

By the proof of Theorem 3.1, $\max_{1 \leq i \leq n} |\hat{\alpha}_i| = O_p[\{(\log n)/T\}^{1/2}]$ and $\|\hat{\boldsymbol{\beta}}\| = o_p(T^{-1/2})$, so that $\hat{f}_i(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) = f_i(0) + \bar{o}_p\{(\log n)/(Th_n)^{1/2}\} = f_i(0) + \bar{o}_p(1)$, which, by condition (A5) (c), implies that $\hat{s}_i = s_i + \bar{o}_p(1)$. Similarly, we can show that $\hat{\boldsymbol{\gamma}}_i =$

$\gamma_i + \bar{o}_p\{(\log n)/(Th_n)^{1/2}\} = \gamma_i + \bar{o}_p(1)$. To prove the second assertion of (A.25), observe that

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

where the second equality is due to Lemma B.2. The third inequality can be deduced from the evaluation analogous to (A.26). Similarly, we can show that $\hat{\Gamma}_n = \Gamma_n + o_p(1)$, so that $\hat{\Gamma}_n^{-1} = \Gamma_n^{-1} + o_p(1)$.

We next show that

$$\hat{\omega}_{ni}^{(1)} = \omega_{ni}^{(1)} + \bar{o}_p(1), \quad \hat{\omega}_{ni}^{(2)} = \omega_{ni}^{(2)} + \bar{o}_p(1), \quad \hat{\omega}_{ni}^{(3)} = \omega_{ni}^{(3)} + \bar{o}_p(1). \quad (\text{A.27})$$

Step 1: proof of $\hat{\omega}_{ni}^{(3)} = \omega_{ni}^{(3)} + \bar{o}_p(1)$. The proof is similar in spirit to the proof of Hahn and Kuersteiner (2011, Lemma 6), but we need a different technique because of the non-differentiability of the indicator function. Observe that

$$\begin{aligned}\omega_{ni}^{(3)} &= \sum_{|j| \leq m_n} \left(1 - \frac{|j|}{T}\right) \text{Cov}\{I(u_{i1} \leq 0), I(u_{i,1+j} \leq 0)\} \\ &\quad + \sum_{m_n+1 \leq |j| \leq T-1} \left(1 - \frac{|j|}{T}\right) \text{Cov}\{I(u_{i1} \leq 0), I(u_{i,1+j} \leq 0)\},\end{aligned}$$

and the latter sum is shown to be $\bar{o}(1)$ by condition (A1) and Lemma C.2. Therefore, it suffices to show that

$$\max_{1 \leq i \leq n} \max_{1 \leq |j| \leq m_n} |\hat{\varrho}_i(j) - \varrho_i(j)| = o(m_n^{-1}). \quad (\text{A.28})$$

By the proof of Theorem 3.1, $\max_{1 \leq i \leq n} |\hat{\alpha}_i| = O_p[\{(\log n)/T\}^{1/2}]$ and $\|\hat{\boldsymbol{\beta}}\| = o_p(T^{-1/2})$, so that by condition (A5),

$$\max_{1 \leq i \leq n} \max_{1 \leq |j| \leq m_n} |\mathbb{E}[I(u_{i1} \leq \alpha + \mathbf{x}'_{i1} \boldsymbol{\beta}) I(u_{i,1+j} \leq \alpha + \mathbf{x}'_{i,1+j} \boldsymbol{\beta})]_{\alpha=\hat{\alpha}_i, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \varrho_i(j)| = O_p[\{(\log n)/T\}^{1/2}].$$

Thus, taking into account that $m_n/T = o(m_n^{-1})$, it suffices to show that

$$\begin{aligned}\max_{1 \leq i \leq n} \max_{1 \leq |j| \leq m_n} \sup_{(\alpha, \boldsymbol{\beta})' \in \mathbb{R}^{p+1}} \left| \frac{1}{T} \sum_{t=\max\{1, -j+1\}}^{\min\{T, T-j\}} \{g_{\alpha, \boldsymbol{\beta}}(u_{it}, \mathbf{x}_{it}; u_{i,t+j}, \mathbf{x}_{i,t+j})\} \right. \\ \left. - \mathbb{E}[g_{\alpha, \boldsymbol{\beta}}(u_{i1}, \mathbf{x}_{i1}; u_{i,1+j}, \mathbf{x}_{i,1+j})] \right| = o_p(m_n^{-1}), \quad (\text{A.29})\end{aligned}$$

where $g_{\alpha, \boldsymbol{\beta}}(u_{it}, \mathbf{x}_{it}; u_{i,t+j}, \mathbf{x}_{i,t+j}) := I(u_{it} \leq \alpha + \mathbf{x}'_{it} \boldsymbol{\beta}) I(u_{i,t+j} \leq \alpha + \mathbf{x}'_{i,t+j} \boldsymbol{\beta})$. We make use of Talagrand's inequality for β -mixing processes (see Appendix C).

Fix any $1 \leq i \leq n$ and $1 \leq |j| \leq m_n$. In what follows, terms const. and $o(\cdot)$ are interpreted as terms independent of i and j . Put $\xi_t := (u_{it}, \mathbf{x}'_{it}, u_{i,t+j}, \mathbf{x}'_{i,t+j})'$. Observe that $\{\xi_t, t = 0, \pm 1, \pm 2, \dots\}$ is β -mixing, and letting $\tilde{\beta}(\cdot)$ denote its β -mixing coefficients, we have $\tilde{\beta}(k) \leq \beta_i(k - |j|)$ for $k > |j|$. Write $g_{\alpha, \beta}(\xi_t) = g_{\alpha, \beta}(u_{it}, \mathbf{x}_{it}; u_{i,t+j}, \mathbf{x}_{i,t+j})$ and define the class of functions $\mathcal{G} = \{g_{\alpha, \beta} : \alpha \in \mathbb{R}, \beta \in \mathbb{R}^p\}$. The class \mathcal{G} is uniformly bounded by 1, and by Lemma 2.6.15 and Theorem 2.6.7 of van der Vaart and Wellner (1996), there exist constants $A \geq 5e$ and $v \geq 1$ (independent of i and j) such that $N(\mathcal{G}, L_1(Q), \epsilon) \leq (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and every probability measure Q on \mathbb{R}^{2p+2} . Take $q = \lceil (T/\log n)^{1/2} \rceil$. By using Lemma 1 of Andrews (1991) and the classical result of Parzen (1957) on the variance evaluation of sample autocovariances, it is shown that $\sup_{(\alpha, \beta)' \in \mathbb{R}^{p+1}} \text{Var}\{\sum_{t=1}^q g_{\alpha, \beta}(\xi_t)/\sqrt{q}\} \leq \text{const.}$. By the exponential β -mixing property (condition (A1)), $r\tilde{\beta}(q) = o(n^{-2})$ with $r = \lceil T/(2q) \rceil$. Therefore, by Propositions C.1 and C.2, with probability at least $1 - o(n^{-2})$ (take $s = 3 \log n$ when applying those propositions),

$$\sup_{(\alpha, \beta)' \in \mathbb{R}^{p+1}} \left| \frac{1}{T} \sum_{t=\max\{1, -j+1\}}^{\min\{T, T-j\}} \{g_{\alpha, \beta}(\xi_t) - \mathbb{E}[g_{\alpha, \beta}(\xi_1)]\} \right| \leq \text{const.} \times \sqrt{(\log n)/T}.$$

The right side is $o(m_n^{-1})$, so that by the union bound, we obtain (A.29).

Step 2: proof of $\hat{\omega}_{ni}^{(1)} = \omega_{ni}^{(1)} + \bar{o}_p(1)$. To begin with, since $\hat{\omega}_{ni}^{(1)} = \bar{O}(1)$ by condition (A9), we have

$$\omega_{ni}^{(1)} = \sum_{1 \leq |j| \leq m_n} \left(1 - \frac{|j|}{T}\right) \{\tau f_i(0) - \phi_i(j)\} + \bar{o}(1),$$

so that

$$|\hat{\omega}_{ni}^{(1)} - \omega_{ni}^{(1)}| \leq m_n |\hat{f}_i - f_i(0)| + m_n \max_{1 \leq |j| \leq m_n} |\hat{\phi}_i(j) - \phi_i(j)| + \bar{o}(1).$$

Because $\hat{f}_i = f_i(0) + \bar{o}\{(\log n)/(Th_n)^{1/2}\}$, the first term on the right side is $\bar{o}(1)$. It remains to show that

$$\max_{1 \leq i \leq n} \max_{1 \leq |j| \leq m_n} |\hat{\phi}_i(j) - \phi_i(j)| = o(m_n^{-1}),$$

which can be shown in a similar way as (A.28). In fact, the problem reduces to showing

$$\begin{aligned} \max_{1 \leq i \leq n} \max_{1 \leq |j| \leq m_n} \sup_{(\alpha, \beta)' \in \mathbb{R}^{p+1}} \left| \frac{1}{Th_n} \sum_{t=\max\{1, -j+1\}}^{\min\{T, T-j\}} \{g_{\alpha, \beta, h_n}(u_{it}, \mathbf{x}_{it}; u_{i,t+j}, \mathbf{x}_{i,t+j}) \right. \\ \left. - \mathbb{E}[g_{\alpha, \beta, h_n}(u_{i1}, \mathbf{x}_{i1}; u_{i,1+j}, \mathbf{x}_{i,1+j})]\} \right| = o_p(m_n^{-1}), \quad (\text{A.30}) \end{aligned}$$

where $g_{\alpha, \beta, h}(u_{it}, \mathbf{x}_{it}; u_{i,t+j}, \mathbf{x}_{i,t+j}) := K((u_{it} - \alpha - \mathbf{x}'_{it}\beta)/h)I(u_{it} \leq \alpha + \mathbf{x}'_{i,t+j}\beta)K((u_{it} - \alpha - \mathbf{x}'_{i,t+j}\beta)/h)I(u_{i,t+j} \leq \alpha + \mathbf{x}'_{i,t+j}\beta)$. Arguing as in the previous case, it is shown

that the left side of (A.30) is $O_p(h_n^{-1}\sqrt{(\log n)/T})$, which is $o_p(m_n^{-1})$ under the present assumption.¹⁸

Step 3: proof of $\hat{\omega}_{ni}^{(2)} = \omega_{ni}^{(2)} + \bar{o}_p(1)$. This step is completely analogous to the previous case.

We have now shown (A.25) and (A.27), so that

$$\begin{aligned}\hat{\mathbf{b}} &= \hat{\Gamma}_n^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \hat{s}_i \left(\hat{\omega}_{ni}^{(1)} \hat{\boldsymbol{\gamma}}_i - \hat{\omega}_{ni}^{(2)} + \frac{\hat{s}_i \hat{\omega}_{ni}^{(3)} \hat{\boldsymbol{\nu}}_i}{2} \right) \right\} \\ &= \Gamma_n^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n s_i \left(\omega_{ni}^{(1)} \boldsymbol{\gamma}_i - \omega_{ni}^{(2)} + \frac{s_i \omega_{ni}^{(3)} \boldsymbol{\nu}_i}{2} \right) \right\} + o_p(1) \\ &\xrightarrow{p} \mathbf{b}.\end{aligned}$$

Therefore, we obtain the desired conclusion. □

B Miscellaneous lemmas

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

Lemma B.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) $\partial_{\alpha_i} H_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) = -f_i(0) + \bar{O}(h_n^r)$;
- (c) $\partial_{\alpha_i}^2 H_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) = -f_i^{(1)}(0) + \bar{O}(h_n^{r-1})$;
- (d) $\partial_{\boldsymbol{\beta}} H_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) = \partial_{\alpha_i} H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = -\mathbb{E}[f_i(0|\mathbf{x}_{it})\mathbf{x}_{i1}] + \bar{O}(h_n^r)$;
- (e) $\partial_{\boldsymbol{\beta}'} H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = -n^{-1} \sum_{i=1}^n \mathbb{E}[f_i(0|\mathbf{x}_{i1})\mathbf{x}_{i1}\mathbf{x}'_{i1}] + O(h_n^r)$;
- (f) $\partial_{\alpha_i}^2 H_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = -\mathbb{E}[f_i^{(1)}(0|\mathbf{x}_{i1})\mathbf{x}_{i1}] + \bar{O}(h_n^{r-1})$;
- (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 $1 \leq j, k, l \leq p$, i.e., 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 $n \rightarrow \infty$ for $1 \leq j \leq p$;

¹⁸We conjecture that the left side of (A.30) is $O_p\{\sqrt{(\log n)/(Th_n)}\}$, but we do not pursue this small improvement here.

(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). \square

Put

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

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

$$\begin{aligned}(a) \quad & \hat{f}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbb{E}[\hat{f}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta})] = o_p\{(\log n)/(Th_n^{2j+1})^{1/2}\}, \quad j = 0, 1; \\ (b) \quad & \tilde{f}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbb{E}[\tilde{f}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta})] = o_p\{(\log n)/(Th_n^{2j+1})^{1/2}\}, \quad j = 1, 2; \\ (c) \quad & \hat{\mathbf{g}}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbb{E}[\hat{\mathbf{g}}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta})] = o_p\{(\log n)/(Th_n^{2j+1})^{1/2}\}, \quad j = 0, 1; \\ (d) \quad & \tilde{\mathbf{g}}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbb{E}[\tilde{\mathbf{g}}_i^{(j)}(\boldsymbol{\alpha}, \boldsymbol{\beta})] = o_p\{(\log n)/(Th_n^{2j+1})^{1/2}\}, \quad j = 1, 2; \\ (e) \quad & \hat{J}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbb{E}[\hat{J}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})] = o_p\{(\log n)/(Th_n)^{1/2}\}; \\ (f) \quad & \tilde{J}_i^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbb{E}[\tilde{J}_i^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta})] = o_p\{(\log n)/(Th_n^3)^{1/2}\}.\end{aligned}$$

As in Horowitz (1998), we employ empirical process theory to prove the lemma, but since the observations are dependent in the time dimension, we use a concentration inequality for β -mixing processes, which is developed in Appendix C. 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 functional $Z(f)$ defined on \mathcal{F} , we use the notation $\|Z(f)\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |Z(f)|$. For a probability measure Q on (S, \mathcal{S}) and a constant $\epsilon > 0$, let $N(\mathcal{F}, L_k(Q), \epsilon)$ denote the ϵ -covering number of \mathcal{F} with respect to the $L_k(Q)$ norm $\|\cdot\|_{L_k(Q)}$, where $k \geq 1$.

Proof of Lemma B.2. We only prove (a) with $j = 0$. The other cases can be shown in a similar way. Put $\xi_{it} := (u_{it}, \mathbf{x}_{it})$, $g_{\alpha, \beta, h}(u, \mathbf{x}) := K((u - \alpha - \mathbf{x}'\beta)/h)$ for $(\alpha, \beta) \in \mathbb{R}^{p+1}$ and $h > 0$, and $\mathcal{G}_{ni} := \{g - \mathbb{E}[g(\xi_{i1})] : g \in \mathcal{G}_n\}$. We apply Propositions C.2 and C.2 to the class \mathcal{G}_{ni} . The pointwise measurability of \mathcal{G}_{ni} is guaranteed by the continuity of $K(\cdot)$. Because of the boundedness of $K(\cdot)$, the class \mathcal{G}_{ni} is uniformly bounded by some constant U (say) independent of i and n . Since condition (A6) guarantees that $K(\cdot)$ is of bounded variation, by Lemma 22 of Nolan and Pollard (1987), there exist constants $A \geq 5e$ and $v \geq 1$ independent of i and n such that $N(\mathcal{G}_{ni}, L_1(Q), U\epsilon) \leq (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and every probability measure Q on \mathbb{R}^{p+1} .

Let $q \in [1, T/2]$. We wish to bound the variance of the sum $\sum_{t=1}^q f(\xi_{it})/\sqrt{q}$. It is standard to show that $\mathbb{E}[g_{\alpha, \beta, h_n}(\xi_{i1})^2] \leq \text{const.} \times h_n$ and $\mathbb{E}[g_{\alpha, \beta, h_n}(\xi_{i1})^2 g_{\alpha, \beta, h_n}(\xi_{i,1+j})^2] \leq \text{const.} \times h_n^2$ where the constants are independent of α, β, i, j and n (the second assertion uses the fact that the joint densities $f_{i,j}(u_1, u_{1+j} | \mathbf{x}_1, \mathbf{x}_{1+j})$ are uniformly bounded, which is assumed in condition (A5) (e)). Thus, by Lemma 1 of Yoshihara (1976) (see Lemma C.2 below), we have $|\text{Cov}(g_{\alpha, \beta, h_n}(\xi_{i1}), g_{\alpha, \beta, h_n}(\xi_{i,1+j}))| \leq \text{const.} \times h_n \beta_i(j)^{1/2}$, which implies that

$$\sup_{g \in \mathcal{G}_{ni}} \text{Var} \left\{ \sum_{t=1}^q g(\xi_{it})/\sqrt{q} \right\} \leq \text{const.} \times \left\{ 1 + 2 \sum_{j=1}^q \beta_i(j)^{1/2} \right\} h_n \leq \text{const.} \times h_n,$$

where the constant on the last expression is independent of i, n and q (recall that condition (A1) ensures that $\sum_{j=1}^{\infty} \sup_{i \geq 1} \beta_i(j)^{1/2} < \infty$). Take $q = q_n = \lfloor (Th_n/\log n)^{1/2} \rfloor$. By the exponential β -mixing property (condition (A1)), $r\beta_i(q) = \bar{o}(n^{-1})$ with $r = \lfloor T/(2q) \rfloor$. Therefore, by Propositions C.1 and C.2, for all $s > 0$, with probability at least $1 - e^{-s} - \bar{o}(n^{-1})$, we have

$$\left\| \sum_{t=1}^T g(\xi_{it}) \right\|_{\mathcal{G}_{ni}} \leq \text{const.} \times \{ \sqrt{Th_n(s \vee \log n)} + s \sqrt{Th_n/\log n} \},$$

where the constant is independent of i, n and s . Taking $s = 2 \log n$ and using the union bound, we obtain the desired result. \square

C Some stochastic inequalities for β -mixing processes

We introduce some stochastic inequalities that can be applied to β -mixing processes. These inequalities are used in the proofs of the theorems above, and are of independent interest.

The first object is to extend Talagrand's (1996) concentration inequality to β -mixing processes. We recall Talagrand's inequality for i.i.d. random variables.

Theorem C.1 (Talagrand (1996); in this form, Massart (2000)). *Let $\xi_i, i = 1, 2, \dots$ be i.i.d. random variables on some measurable space S . Let \mathcal{F} a pointwise measurable class of functions on S such that $\mathbb{E}[f(\xi_1)] = 0$ and $\sup_{x \in S} |f(x)| \leq U$ (see van der Vaart and Wellner, 1996, Section 2.3 for pointwise measurability). Let σ^2 be any positive constant such that $\sigma^2 \geq \sup_{f \in \mathcal{F}} \mathbb{E}[f(\xi_1)^2]$. Put $Z := \sup_{f \in \mathcal{F}} |\sum_{i=1}^n f(\xi_i)|$. Then, for all $s > 0$, we have*

$$\mathbb{P}\{Z \geq 2\mathbb{E}[Z] + C\sigma\sqrt{ns} + sCU\} \leq e^{-s},$$

where C is a universal constant.

In what follows, let $\{\xi_t, t \in \mathbb{Z}\}$ be a stationary process taking values in a measurable space (S, \mathcal{S}) . We assume that S is a Polish space and \mathcal{S} is its Borel σ -field. For a function f on S and a positive integer q , define $\sigma_q^2(f) := \text{Var}\{f(\xi_1)\} + 2 \sum_{j=1}^{q-1} (1 - j/q) \text{Cov}\{f(\xi_1), f(\xi_{1+j})\}$, which is the variance of the sum $\sum_{t=1}^q f(\xi_t)/\sqrt{q}$. Let $\beta(j)$ denote the β -mixing coefficient of $\{\xi_t\}$.

The main technical device we use is a blocking technique developed in Yu (1993, 1994) and Arcones and Yu (1994). The blocking technique enables us to employ maximal inequalities and the symmetrization technique available in the i.i.d. case. We introduce some notation and a key lemme related to the blocking technique. Divide the T -sequence $\{1, \dots, T\}$ into blocks of length q with $q \in [1, T/2]$, one after the other:

$$\begin{aligned} H_k &= \{t : 2(k-1)q + 1 \leq t \leq (2k-1)q\}, \\ T_k &= \{t : (2k-1)q + 1 \leq t \leq 2kq\}, \end{aligned}$$

where $k = 1, \dots, r := \lceil T/(2q) \rceil$. Put $\Xi_k := (\xi_t, t \in H_k)$. With a slight abuse of notation, for a function $f : S \rightarrow \mathbb{R}$, we write $f(\Xi_k) = \sum_{t \in H_k} f(\xi_t)$. Let $\tilde{\Xi}_k = (\tilde{\xi}_t, t \in H_k)$, $k = 1, \dots, r$ be independent blocks such that each $\tilde{\Xi}_k$ has the same distribution as Ξ_1 . The next lemma, which is a key to the blocking technique, is due to Eberlain (1984).

Lemma C.1. *Work with the same notation as above. For every Borel measurable subset A of S^{qr} , we have*

$$\left| \mathbb{P}\{(\Xi_1, \dots, \Xi_r) \in A\} - \mathbb{P}\{(\tilde{\Xi}_1, \dots, \tilde{\Xi}_r) \in A\} \right| \leq r\beta(q).$$

Suppose that we have a pointwise measurable class of functions \mathcal{F} on S such that $\sup_{x \in S} |f(x)| \leq U$ for some constant U . For each $f \in \mathcal{F}$, observe that

$$\left| \sum_{t=1}^T f(\xi_t) \right| \leq \left| \sum_{k=1}^r \sum_{t \in H_k} f(\xi_t) \right| + \left| \sum_{k=1}^r \sum_{t \in T_k} f(\xi_t) \right| + (T - 2qr)U.$$

Since the first and second terms on the right side have the same distribution, for all $s > 0$, we have

$$\begin{aligned}
& \mathbb{P} \left\{ \left\| \sum_{t=1}^T f(\xi_t) \right\|_{\mathcal{F}} \geq 2s + (T - 2qr)U \right\} \\
& \leq \mathbb{P} \left\{ \left\| \sum_{k=1}^r \sum_{t \in H_k} f(\xi_t) \right\|_{\mathcal{F}} + \left\| \sum_{k=1}^r \sum_{t \in T_k} f(\xi_t) \right\|_{\mathcal{F}} \geq 2s \right\} \\
& \leq \mathbb{P} \left\{ \left\| \sum_{k=1}^r \sum_{t \in H_k} f(\xi_t) \right\|_{\mathcal{F}} \geq s \right\} + \mathbb{P} \left\{ \left\| \sum_{k=1}^r \sum_{t \in T_k} f(\xi_t) \right\|_{\mathcal{F}} \geq s \right\} \\
& = 2\mathbb{P} \left\{ \left\| \sum_{k=1}^r \sum_{t \in H_k} f(\xi_t) \right\|_{\mathcal{F}} \geq s \right\} \\
& = 2\mathbb{P} \left\{ \left\| \sum_{k=1}^r f(\Xi_k) \right\|_{\mathcal{F}} \geq s \right\}
\end{aligned}$$

By Lemma C.1, the last expression is bounded by

$$2\mathbb{P} \left\{ \left\| \sum_{k=1}^r f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \geq s \right\} + 2r\beta(q).$$

Recall that $\tilde{\Xi}_k$, $k = 1, \dots, r$ are i.i.d. blocks. By Talagrand's inequality, for all $s > 0$, we have

$$\mathbb{P} \left\{ \left\| \sum_{k=1}^r f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \geq 2\mathbb{E} \left[\left\| \sum_{k=1}^r f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \right] + C\sigma_q\sqrt{sT} + sqCU \right\} \leq e^{-s}$$

where σ_q^2 is any positive constant such that $\sup_{f \in \mathcal{F}} \sigma_q^2(f) \leq \sigma_q^2 \leq 2qU^2$. Therefore, we obtain the next proposition:

Proposition C.1. *Work with the same notation as above. Then, for all $s > 0$, we have*

$$\mathbb{P} \left\{ \left\| \sum_{t=1}^T f(\xi_t) \right\|_{\mathcal{F}} \geq 4\mathbb{E} \left[\left\| \sum_{k=1}^r f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \right] + C\sigma_q\sqrt{sT} + sqCU \right\} \leq 2e^{-s} + 2r\beta(q).$$

For the evaluation of the term $\mathbb{E}[\|\sum_{k=1}^r f(\tilde{\Xi}_k)\|_{\mathcal{F}}]$, the next proposition is useful. Recall that for a probability measure Q on (S, \mathcal{S}) and a constant $\epsilon > 0$, let $N(\mathcal{F}, L_k(Q), \epsilon)$ denote the ϵ -covering number of \mathcal{F} with respect to the $L_k(Q)$ norm $\|\cdot\|_{L_k(Q)}$, where $k \geq 1$.

Proposition C.2. *Work with same notation as above. Assume that there exist constants $A \geq 5e$ and $v \geq 1$ such that $N(\mathcal{F}, L_1(Q), U\epsilon) \leq (A/\epsilon)^v$ for every $0 < \epsilon < 1$ and*

every probability measure Q on S . Then, we have

$$\mathbb{E} \left[\left\| \sum_{k=1}^r f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \right] \leq C \left[qvU \log \frac{\sqrt{q}A'U}{\sigma_q} + \sqrt{v}\sqrt{T}\sigma_q \sqrt{\log \frac{\sqrt{q}A'U}{\sigma_q}} \right], \quad (\text{C.1})$$

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

Proof. The proof is based on the proof of Gine and Guillou (2001, Proposition 2.1), but some modifications are needed. Recall that $\tilde{\Xi}_k, k = 1, \dots, r$ are i.i.d. Let $\epsilon_1, \dots, \epsilon_r$ be i.i.d. Rademacher random variables independent of $\tilde{\Xi}_1, \dots, \tilde{\Xi}_r$. By Lemma 2.3.1 of van der Vaart and Wellner (1996), we have

$$\begin{aligned} \mathbb{E} \left[\left\| \sum_{k=1}^r f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \right] &\leq 2\mathbb{E} \left[\left\| \sum_{k=1}^r \epsilon_k f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \right] \\ &= 2qU\mathbb{E} \left[\left\| \sum_{k=1}^r \epsilon_k \varphi(\tilde{\Xi}_k) \right\|_{\mathcal{H}} \right], \end{aligned} \quad (\text{C.2})$$

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

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

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

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

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

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

Integration by parts gives

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

from which we have

$$\int_a^\infty \frac{\sqrt{\log \epsilon}}{\epsilon^2} d\epsilon \leq \frac{2\sqrt{\log a}}{a}, \quad a \geq e.$$

Therefore, we have

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

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

Now, by Corollary 3.4 of Talagrand (1994),

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

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

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

Put

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

Then, Z satisfies

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

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

$$\begin{aligned}
Z &\leq 4Cv \log \frac{\sqrt{q}A'U}{\sigma_q} + \sqrt{16C^2v^2 \left(\log \frac{\sqrt{q}A'U}{\sigma_q} \right)^2 + C \frac{vr\sigma_q^2}{qU^2} \log \frac{\sqrt{q}A'U}{\sigma_q}} \\
&\leq 8Cv \log \frac{\sqrt{q}A'U}{\sigma_q} + \sqrt{C} \frac{\sqrt{v}\sqrt{r}\sigma_q}{\sqrt{q}U} \sqrt{\log \frac{\sqrt{q}A'U}{\sigma_q}} \\
&\leq C' \left[v \log \frac{\sqrt{q}A'U}{\sigma_q} + \frac{\sqrt{v}\sqrt{r}\sigma_q}{\sqrt{q}U} \sqrt{\log \frac{\sqrt{q}A'U}{\sigma_q}} \right], \tag{C.3}
\end{aligned}$$

where the second inequality is due to $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b > 0$, and C' is another universal constant. Combining (C.2) and (C.3) yields the desired inequality. \square

In the case that \mathcal{F} is singleton, i.e., $\mathcal{F} = \{f\}$, Proposition C.2 gives a Bernstein type inequality for β -mixing processes. In that case,

$$\begin{aligned}
\mathbb{E} \left[\left\| \sum_{k=1}^r f(\tilde{\Xi}_k) \right\|_{\mathcal{F}} \right] &= \mathbb{E} \left[\left| \sum_{k=1}^r f(\tilde{\Xi}_k) \right| \right] \\
&\leq \sqrt{\mathbb{E} \left[\left| \sum_{k=1}^r f(\tilde{\Xi}_k) \right|^2 \right]} \\
&\leq \sqrt{rq} \sigma_q(f).
\end{aligned}$$

Although such inequalities are known in the literature (cf. Fan and Yao, 2003, Theorem 2.18), we present the special case of \mathcal{F} being singleton as a corollary, since its form is more convenient in the present context.

Corollary C.1. *Let f be a function on S such that $\sup_{x \in S} |f(x)| \leq U$ and $\mathbb{E}[f(\xi_1)] = 0$. Then, for all $s > 0$, we have*

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

where C is a universal constant.

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

Lemma C.2. *Work with the same notation as above. Let j be a fixed positive integer. Let f and g be functions on S such that $\mathbb{E}[f(\xi_1)] = \mathbb{E}[g(\xi_{1+j})] = 0$, and for some positive constants δ and M ,*

$$\mathbb{E}[|f(\xi_1)|^{1+\delta}] \mathbb{E}[|g(\xi_{1+j})|^{1+\delta}] \leq M, \quad \mathbb{E}[|f(\xi_1)g(\xi_{1+j})|^{1+\delta}] \leq M. \tag{C.4}$$

Then, we have

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

A direct consequence of Lemma C.2 is that if there exist positive constants δ and M such that (C.4) holds for any positive integer j and $\sum_{j=1}^{\infty} \beta(j)^{1/(1+\delta)} < \infty$, then the infinite sum $\sum_{j=1}^{\infty} \text{Cov}\{f(\xi_1), g(\xi_{1+j})\}$ is absolutely convergent, and in particular, for any positive integer q ,

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

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

References

- ABREVAYA, J., AND C. M. DAHL (2008): “The Effects of Birth Inputs on Birth-weight: 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.
- ANDREWS, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–858.
- 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.
- ARCONES, M. A., AND B. YU (1994): “Central Limit Theorems for Empirical and U-processes of Stationary Mixing Sequences,” *Journal of Theoretical Probabilities*, 7, 47–71.
- 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 Jackknife Estimation of Fixed-Effect Models,” Mimeo.
- EBERLAIN, E. (1984): “Weak Convergence of Partial Sums of Absolutely Regular Sequences,” *Statistics and Probability Letters*, 2, 291–293.
- FAN, J., AND Q. YAO (2003): *Nonlinear Time Series*. Springer-Verlag.
- FERNANDEZ-VAL, I. (2005): “Bias Correction in Panel Data Models with Individual Specific Parameters,” Mimeo.
- GAMPER-RABINDRAN, S., S. KHAN, AND C. TIMMINS (2010): “The Impact of Piped Water Provision on Infant Mortality in Brazil: a Quantile Panel Data Approach,” *Journal of Development Economics*, 92, 188–200.
- GINE, E., AND A. GUILLOU (2001): “On Consistency of Kernel Density Estimators for Randomly Censored Data: Rates Holding Uniformly over Adaptive Intervals,” *Annales de l’Institut Henri Poincaré*, 37, 503–522.
- 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 (2011): “Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When both n and T are Large,” *Econometric Theory*, forthcoming.
- 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 (2011): “Asymptotics for Panel Quantile Regression Models with Individual 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.
- LIPSITZ, S. R., G. M. FITZMAURICE, G. MOLENBERGHS, AND L. P. ZHAO (1997): “Quantile Regression Methods for Longitudinal Data with Drop-outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus,” *Journal of the Royal Statistical Society, Series C*, 46, 463–476.
- MASSART, P. (2000): “About the Constants in Talagrand’s Concentration Inequalities for Empirical Processes,” *Annals of Probability*, 28, 863–884.
- 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.
- PARZEN, E. (1957): “On Consistent Estimates of the Spectrum of a Stationary Time Series,” *Annals of Mathematical Statistics*, 28, 329–348.
- PHILLIPS, P. C. (1991): “A Shortcut to LAD Estimator Asymptotics,” *Econometric Theory*, 7, 450–463.
- 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.
- SILVA, D. G. D., G. KOSMOPOULOU, AND C. LAMARCHE (2009): “The Effect of Information on the Bidding and Survival of Entrants in Procurement Auctions,” *Journal of Public Economics*, 93, 56–72.
- TALAGRAND, M. (1994): “Sharper Bounds for Gaussian and Empirical Processes,” *Annals of Probability*, 22, 28–76.
- (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.
- WANG, H. J., AND M. FYGENSON (2009): “Inference for Censored Quantile Regression Models in Longitudinal Studies,” *Annals of Statistics*, 37, 756–781.
- WEI, Y., AND X. HE (2006): “Conditional Growth Charts,” *Annals of Statistics*, 34, 2069–2097.
- WOUTERSEN, T. M. (2002): “Robustness Against Incidental Parameters,” Mimeo, University of Western Ontario.
- YOSHIHARA, K. (1976): “Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35, 237–252.

- (1978): “Moment inequalities for Mixing Sequences,” *Kodai Mathematical Journal*, 1, 316–328.
- YU, B. (1993): “Density Estimation in the L^∞ norm for Dependent Data with Applications to the Gibbs Sampler,” *Annals of Statistics*, 21, 711–735.
- (1994): “Rates of Convergence for Empirical Processes of Stationary Mixing Sequences,” *Annals of Probability*, 22, 94–116.

Table 3: Bias, Standard Deviation, and Root Mean Squared Error. Standard Normal innovation model. $\theta = 1$.

BIAS	τ	$\hat{\beta}_{\text{KB}}$						$\hat{\beta}$						$\hat{\beta}^1$						$\hat{\beta}_{1/2}$						$\hat{\beta}_{\text{KB}}^1$					
		T = 10		T = 20		T = 50		T = 10		T = 20		T = 50		T = 10		T = 20		T = 50		T = 10		T = 20		T = 50		T = 10		T = 20		T = 50	
		0.25	0.50	0.75	0.060	0.064	-0.002	-0.063	0.042	-0.002	-0.045	0.067	0.009	-0.047	0.044	0.005	-0.042	0.067	0.009	-0.045	0.140	0.128	0.138	0.189	0.180	0.189	0.180	0.189	0.180	0.189	0.180
	$n = 50$	0.067	0.004	-0.060	0.064	-0.002	-0.063	0.042	-0.002	-0.045	0.067	0.009	-0.047	0.044	0.005	-0.042	0.067	0.009	-0.045	0.140	0.128	0.138	0.189	0.180	0.189	0.180	0.189	0.180	0.189	0.180	
	$n = 100$	0.059	0.000	-0.066	0.058	-0.002	-0.066	0.035	-0.004	-0.046	0.059	0.001	-0.064	0.035	-0.003	-0.046	0.059	0.001	-0.064	0.093	0.089	0.093	0.126	0.113	0.126	0.113	0.126	0.113	0.126	0.113	
	$n = 500$	0.064	0.000	-0.062	0.064	0.000	-0.062	0.035	-0.003	-0.040	0.064	0.000	-0.062	0.035	-0.003	-0.040	0.064	0.000	-0.062	0.045	0.040	0.042	0.059	0.051	0.059	0.051	0.059	0.051	0.059	0.051	
	$n = 50$	0.035	0.008	-0.028	0.022	0.005	-0.028	0.003	0.004	-0.011	0.030	0.013	-0.018	0.016	0.007	-0.012	0.030	0.013	-0.018	0.103	0.092	0.098	0.114	0.100	0.114	0.100	0.109	0.100	0.109	0.100	
	$n = 100$	0.026	0.000	-0.026	0.023	0.000	-0.025	0.002	0.000	-0.009	0.027	0.001	-0.025	0.002	0.000	-0.010	0.027	0.001	-0.025	0.069	0.067	0.071	0.078	0.074	0.078	0.074	0.079	0.078	0.074	0.079	
	$n = 500$	0.029	0.001	-0.030	0.029	0.001	-0.030	0.012	0.000	-0.015	0.029	0.001	-0.030	0.012	0.000	-0.015	0.029	0.001	-0.030	0.031	0.029	0.031	0.035	0.033	0.035	0.033	0.035	0.033	0.035	0.033	
	$n = 50$	0.011	-0.001	-0.012	0.010	-0.002	-0.014	0.001	-0.002	-0.006	0.013	0.003	-0.007	0.003	-0.002	-0.004	0.013	0.003	-0.007	0.063	0.058	0.063	0.064	0.058	0.064	0.058	0.064	0.058	0.064	0.058	
	$n = 100$	0.012	0.001	-0.011	0.012	0.001	-0.011	0.003	0.001	-0.002	0.013	0.001	-0.010	0.003	0.001	-0.002	0.013	0.001	-0.010	0.044	0.040	0.042	0.044	0.041	0.044	0.041	0.044	0.041	0.044	0.041	
	$n = 500$	0.005	0.000	-0.004	0.001	-0.001	-0.005	0.003	0.001	-0.002	0.005	0.002	-0.002	0.003	0.001	-0.002	0.005	0.002	-0.002	0.020	0.018	0.019	0.020	0.018	0.020	0.018	0.020	0.018	0.020	0.018	
	$n = 100$	0.006	-0.001	-0.006	0.005	-0.001	-0.006	0.000	-0.001	-0.002	0.006	-0.001	-0.002	0.000	-0.001	-0.002	0.006	-0.001	-0.002	0.044	0.040	0.042	0.044	0.040	0.044	0.040	0.044	0.040	0.044	0.040	
	$n = 500$	0.005	0.000	-0.005	0.005	0.000	-0.005	0.001	0.000	-0.001	0.005	0.000	-0.001	0.001	0.000	-0.001	0.005	0.000	-0.001	0.031	0.028	0.031	0.031	0.028	0.031	0.028	0.031	0.028	0.031	0.028	
SD	$T = 10$	0.137	0.126	0.137	0.138	0.124	0.136	0.189	0.178	0.181	0.140	0.128	0.138	0.189	0.180	0.181	0.140	0.128	0.181	0.189	0.178	0.181	0.189	0.180	0.189	0.180	0.189	0.180	0.189	0.180	
	$n = 100$	0.093	0.088	0.093	0.093	0.088	0.093	0.126	0.113	0.124	0.093	0.089	0.093	0.126	0.113	0.124	0.093	0.089	0.093	0.093	0.089	0.093	0.126	0.113	0.126	0.113	0.126	0.113	0.126	0.113	
	$n = 500$	0.045	0.040	0.042	0.045	0.040	0.042	0.059	0.051	0.055	0.045	0.040	0.042	0.059	0.051	0.055	0.045	0.040	0.042	0.045	0.040	0.042	0.059	0.051	0.059	0.051	0.059	0.051	0.059	0.051	
	$n = 50$	0.101	0.090	0.096	0.100	0.090	0.096	0.113	0.099	0.109	0.103	0.092	0.098	0.113	0.100	0.109	0.103	0.092	0.098	0.069	0.067	0.071	0.078	0.074	0.078	0.074	0.079	0.078	0.074	0.079	
	$n = 100$	0.068	0.067	0.070	0.068	0.066	0.070	0.078	0.074	0.079	0.078	0.074	0.079	0.078	0.074	0.079	0.078	0.074	0.079	0.031	0.029	0.031	0.035	0.033	0.035	0.033	0.035	0.033	0.035	0.033	
	$n = 500$	0.031	0.029	0.031	0.031	0.029	0.031	0.035	0.033	0.035	0.031	0.029	0.031	0.035	0.033	0.035	0.031	0.029	0.031	0.063	0.057	0.063	0.064	0.057	0.064	0.057	0.064	0.057	0.064	0.057	
	$n = 50$	0.061	0.056	0.061	0.061	0.056	0.062	0.064	0.057	0.064	0.061	0.056	0.062	0.064	0.057	0.064	0.061	0.056	0.062	0.044	0.041	0.044	0.044	0.041	0.044	0.041	0.044	0.041	0.044	0.041	
	$n = 100$	0.042	0.040	0.042	0.042	0.040	0.042	0.044	0.041	0.044	0.042	0.040	0.042	0.044	0.041	0.044	0.042	0.040	0.042	0.020	0.018	0.020	0.020	0.018	0.020	0.018	0.020	0.018	0.020	0.018	
	$n = 500$	0.019	0.018	0.019	0.019	0.018	0.019	0.020	0.018	0.020	0.019	0.018	0.020	0.020	0.018	0.020	0.019	0.018	0.020	0.044	0.040	0.042	0.044	0.040	0.044	0.040	0.044	0.040	0.044	0.040	
	$n = 50$	0.043	0.039	0.041	0.043	0.039	0.041	0.044	0.039	0.042	0.043	0.039	0.042	0.044	0.039	0.042	0.043	0.039	0.042	0.020	0.018	0.020	0.020	0.018	0.020	0.018	0.020	0.018	0.020	0.018	
	$n = 100$	0.030	0.028	0.030	0.030	0.028	0.030	0.031	0.028	0.031	0.030	0.028	0.031	0.031	0.028	0.031	0.030	0.028	0.031	0.044	0.040	0.042	0.044	0.040	0.044	0.040	0.044	0.040	0.044	0.040	
	$n = 500$	0.014	0.013	0.013	0.014	0.013	0.013	0.014	0.013	0.014	0.014	0.013	0.014	0.014	0.013	0.014	0.014	0.013	0.014	0.031	0.028	0.030	0.031	0.028	0.031	0.028	0.031	0.028	0.031	0.028	
RMSE	$T = 10$	0.153	0.126	0.149	0.152	0.124	0.150	0.193	0.178	0.187	0.155	0.128	0.146	0.194	0.180	0.186	0.155	0.128	0.187	0.194	0.178	0.187	0.194	0.180	0.194	0.180	0.194	0.180	0.194	0.180	
	$n = 100$	0.110	0.088	0.114	0.110	0.088	0.114	0.131	0.113	0.132	0.111	0.089	0.113	0.131	0.113	0.132	0.111	0.089	0.113	0.093	0.089	0.093	0.131	0.113	0.131	0.113	0.131	0.113	0.131	0.113	
	$n = 500$	0.078	0.040	0.074	0.078	0.040	0.074	0.071	0.051	0.068	0.078	0.040	0.074	0.071	0.051	0.068	0.078	0.040	0.074	0.078	0.040	0.074	0.071	0.051	0.078	0.040	0.074	0.071	0.051	0.068	
	$n = 50$	0.106	0.091	0.100	0.102	0.090	0.100	0.113	0.099	0.110	0.107	0.093	0.110	0.113	0.099	0.110	0.107	0.093	0.110	0.078	0.074	0.075	0.078	0.074	0.079	0.078	0.074	0.079	0.078	0.074	
	$n = 100$	0.073	0.067	0.075	0.072	0.066	0.075	0.078	0.074	0.079	0.074	0.067	0.075	0.078	0.074	0.079	0.074	0.067	0.075	0.036	0.033	0.036	0.036	0.033	0.036	0.033	0.036	0.033	0.036	0.033	
	$n = 500$	0.042	0.029	0.043	0.042	0.029	0.043	0.036	0.033	0.038	0.042	0.029	0.043	0.036	0.033	0.038	0.042	0.029	0.043	0.064	0.058	0.064	0.064	0.058	0.064	0.058	0.064	0.058	0.064	0.058	
	$n = 50$	0.062	0.056	0.063	0.062	0.056	0.063	0.064	0.058	0.065	0.064	0.058	0.065	0.064	0.058	0.065	0.064	0.058	0.065	0.020	0.018	0.022	0.020	0.018	0.022	0.018	0.022	0.018	0.022	0.018	
	$n = 100$	0.044	0.040	0.043	0.044	0.040	0.043	0.045	0.041	0.044	0.044	0.040	0.044	0.045	0.041	0.044	0.044	0.040	0.044	0.044	0.040	0.044	0.044	0.040	0.044	0.040	0.044	0.040	0.044	0.040	
	$n = 500$	0.023	0.018	0.022	0.023	0.018	0.022	0.020	0.018	0.020	0.023	0.018	0.022	0.020	0.018	0.022	0.023	0.018	0.022	0.044	0.040	0.042	0.044	0.040	0.044	0.040	0.044	0.040	0.044	0.040	
	$n = 50$	0.043	0.039	0.041	0.043	0.039	0.041	0.044	0.039	0.042	0.043	0.039	0.042	0.044	0.039	0.042	0.043	0.039	0.042	0.031	0.028	0.031	0.031	0.028	0.031	0.028	0.031	0.028	0.031	0.028	
	$n = 100$	0.031	0.028	0.031	0.031	0.028	0.031	0.031	0.028	0.031	0.031	0.028	0.031	0.031	0.028</																

Table 4: Bias, Standard Deviation, and Root Mean Squared Error. Chi-square innovation model. $\theta = 1$.

BIAS	T	τ	$\hat{\beta}_{\text{KB}}$						$\hat{\beta}$						$\hat{\beta}_{1/2}$						$\hat{\beta}_{\text{KB}}^1$									
			0.25		0.50		0.75		0.25		0.50		0.75		0.25		0.50		0.75		0.25		0.50		0.75					
			0.062	-0.056	-0.267	0.054	-0.054	-0.273	-0.028	-0.051	-0.227	0.059	-0.045	-0.224	-0.020	-0.054	-0.221	0.051	-0.064	-0.261	0.050	-0.061	-0.262	-0.053	-0.062	-0.242	-0.052	-0.065	-0.242	
		$n = 50$	0.051	-0.064	-0.261	0.050	-0.061	-0.262	0.050	-0.056	-0.262	0.022	-0.027	-0.121	-0.030	-0.042	-0.051	0.024	-0.015	-0.094	0.015	-0.035	-0.126	-0.041	-0.019	-0.044	0.024	-0.030	-0.049	-0.227
		$n = 100$	0.020	-0.028	-0.123	0.022	-0.027	-0.121	0.017	-0.034	-0.126	0.016	-0.034	-0.130	-0.035	-0.014	-0.046	0.015	-0.035	-0.126	0.016	-0.034	-0.130	-0.035	-0.014	-0.046	0.015	-0.030	-0.049	-0.227
		$n = 500$	0.015	-0.036	-0.129	0.016	-0.034	-0.130	0.006	-0.012	-0.051	0.006	-0.012	-0.051	-0.010	0.001	-0.009	0.006	-0.012	-0.051	0.006	-0.012	-0.051	-0.010	0.001	-0.009	0.006	-0.011	-0.006	-0.011
		$n = 100$	0.016	-0.034	-0.130	0.016	-0.034	-0.130	0.004	-0.017	-0.054	0.004	-0.017	-0.054	-0.010	0.001	-0.009	0.004	-0.017	-0.054	0.004	-0.017	-0.054	-0.010	0.001	-0.009	0.004	-0.018	-0.006	-0.023
		$n = 500$	-0.002	-0.019	-0.067	-0.008	-0.021	-0.073	-0.003	-0.016	-0.036	-0.003	-0.016	-0.036	-0.010	-0.009	-0.015	-0.002	-0.013	-0.033	-0.004	-0.006	-0.026	-0.002	-0.007	-0.026	-0.002	-0.006	-0.012	-0.008
		$n = 100$	0.005	-0.019	-0.054	0.004	-0.017	-0.054	0.004	-0.017	-0.054	0.004	-0.017	-0.054	-0.002	0.001	-0.002	0.004	-0.007	-0.026	0.004	-0.006	-0.026	-0.002	0.001	-0.002	0.004	-0.002	0.000	-0.003
		$n = 500$	0.006	-0.012	-0.051	0.006	-0.012	-0.051	0.006	-0.012	-0.051	0.006	-0.012	-0.051	-0.002	0.001	-0.002	0.004	-0.007	-0.026	0.006	-0.012	-0.051	-0.002	0.001	-0.002	0.004	-0.002	0.000	-0.003
		$n = 100$	0.004	-0.013	-0.033	-0.003	-0.016	-0.036	0.004	-0.006	-0.025	0.004	-0.006	-0.025	-0.004	0.000	-0.003	0.004	-0.007	-0.026	0.004	-0.006	-0.026	-0.004	0.000	-0.003	0.004	-0.004	0.000	-0.003
		$n = 500$	0.225	0.283	0.381	0.224	0.281	0.381	0.161	0.201	0.261	0.161	0.201	0.261	0.429	0.482	0.710	0.230	0.291	0.390	0.429	0.482	0.710	0.230	0.291	0.390	0.429	0.482	0.711	0.482
		$n = 100$	0.161	0.202	0.261	0.072	0.089	0.117	0.072	0.089	0.117	0.174	0.204	0.279	0.313	0.353	0.530	0.161	0.202	0.262	0.161	0.202	0.262	0.313	0.353	0.530	0.161	0.202	0.262	0.313
		$n = 500$	0.072	0.089	0.117	0.175	0.204	0.279	0.174	0.204	0.279	0.121	0.147	0.199	0.139	0.158	0.231	0.072	0.089	0.117	0.139	0.158	0.231	0.139	0.158	0.231	0.139	0.158	0.231	0.231
		$n = 100$	0.175	0.204	0.279	0.121	0.147	0.199	0.121	0.147	0.199	0.052	0.061	0.086	0.071	0.079	0.127	0.121	0.147	0.199	0.121	0.147	0.199	0.162	0.193	0.295	0.162	0.193	0.295	0.295
		$n = 500$	0.052	0.061	0.086	0.101	0.126	0.169	0.052	0.061	0.086	0.052	0.061	0.086	0.071	0.079	0.127	0.052	0.061	0.086	0.071	0.079	0.127	0.052	0.061	0.086	0.071	0.079	0.127	0.127
		$n = 100$	0.101	0.126	0.169	0.073	0.091	0.125	0.073	0.091	0.125	0.111	0.139	0.197	0.111	0.139	0.197	0.111	0.139	0.197	0.111	0.139	0.197	0.111	0.139	0.197	0.111	0.139	0.197	0.197
		$n = 500$	0.073	0.091	0.125	0.032	0.041	0.056	0.032	0.041	0.056	0.081	0.100	0.146	0.081	0.100	0.146	0.073	0.091	0.125	0.081	0.100	0.146	0.081	0.100	0.146	0.081	0.100	0.146	0.146
		$n = 100$	0.032	0.041	0.056	0.073	0.089	0.126	0.032	0.041	0.056	0.036	0.045	0.065	0.036	0.045	0.065	0.032	0.041	0.056	0.036	0.045	0.065	0.036	0.045	0.065	0.036	0.045	0.065	0.065
		$n = 500$	0.073	0.088	0.126	0.052	0.063	0.087	0.052	0.063	0.087	0.077	0.092	0.135	0.077	0.092	0.135	0.077	0.092	0.135	0.077	0.092	0.135	0.077	0.092	0.135	0.077	0.092	0.135	0.135
		$n = 100$	0.052	0.063	0.087	0.024	0.029	0.041	0.024	0.029	0.041	0.054	0.066	0.093	0.054	0.066	0.093	0.052	0.063	0.087	0.052	0.063	0.087	0.052	0.063	0.087	0.052	0.063	0.087	0.087
		$n = 500$	0.024	0.029	0.041	0.233	0.288	0.466	0.231	0.286	0.468	0.231	0.286	0.468	0.430	0.485	0.745	0.237	0.295	0.450	0.430	0.486	0.744	0.237	0.295	0.450	0.430	0.486	0.744	0.744
		$n = 100$	0.169	0.212	0.369	0.088	0.104	0.287	0.168	0.210	0.370	0.088	0.104	0.287	0.317	0.358	0.583	0.169	0.212	0.367	0.169	0.212	0.367	0.317	0.358	0.583	0.169	0.212	0.367	0.583
		$n = 500$	0.088	0.104	0.287	0.176	0.206	0.305	0.088	0.104	0.287	0.175	0.205	0.304	0.142	0.165	0.324	0.088	0.104	0.287	0.142	0.165	0.324	0.142	0.165	0.324	0.142	0.165	0.324	0.324
		$n = 100$	0.176	0.206	0.305	0.122	0.152	0.238	0.175	0.205	0.304	0.122	0.151	0.236	0.234	0.255	0.425	0.182	0.211	0.306	0.182	0.211	0.306	0.182	0.211	0.306	0.182	0.211	0.306	0.306
		$n = 500$	0.122	0.152	0.238	0.054	0.070	0.156	0.122	0.151	0.236	0.054	0.070	0.156	0.167	0.194	0.298	0.123	0.152	0.236	0.167	0.194	0.298	0.123	0.152	0.236	0.167	0.194	0.298	0.298
		$n = 100$	0.054	0.070	0.156	0.101	0.127	0.182	0.054	0.070	0.156	0.101	0.128	0.184	0.079	0.080	0.135	0.054	0.070	0.156	0.079	0.080	0.135	0.054	0.070	0.156	0.079	0.080	0.135	0.135
		$n = 500$	0.101	0.127	0.182	0.073	0.093	0.136	0.101	0.128	0.184	0.073	0.092	0.136	0.114	0.140	0.199	0.105	0.131	0.179	0.114	0.140	0.199	0.105	0.131	0.179	0.114	0.140	0.199	0.199
		$n = 100$	0.073	0.093	0.136	0.032	0.043	0.076	0.073	0.092	0.136	0.032	0.043	0.076	0.082	0.100	0.146	0.074	0.092	0.135	0.082	0.100	0.146	0.074	0.092	0.135	0.082	0.100	0.146	0.146
		$n = 500$	0.032	0.043	0.076	0.073	0.089	0.131	0.032	0.043	0.076	0.073	0.089	0.131	0.037	0.045	0.066	0.032	0.043	0.066	0.037	0.045	0.066	0.032	0.043	0.066	0.037	0.045	0.066	0.066
		$n = 100$	0.073	0.089	0.131	0.052	0.063	0.090	0.073	0.089	0.131	0.052	0.063	0.090	0.077	0.093	0.136	0.075	0.092	0.131	0.077	0.093	0.136	0.075	0.092	0.131	0.077	0.093	0.136	0.136
		$n = 500$	0.052	0.063	0.090	0.024	0.030	0.048	0.052	0.063	0.090	0.024	0.030	0.048	0.054	0.066	0.093	0.052	0.063	0.090	0.054	0.066	0.093	0.052	0.063	0.090	0.054	0.066	0.093	0.093
		$n = 100$	0.024	0.030	0.048	0.024	0.030	0.048	0.024	0.030	0.048	0.024	0.030	0.048	0.025	0.031	0.044	0.024	0.030	0.044	0.025	0.031	0.044	0.024	0.030	0.044	0.025	0.031	0.044	0.044

Notes: Monte Carlo experiments based on 1000 repetitions.