Stat Papers (2017) 58:95–122
DOI 10.1007/s00362-015-0689-8

REGULAR ARTICLE

Jackknife empirical likelihood of error variance in partially linear varying-coefficient errors-in-variables models

Ai-Ai Liu[1] · **Han-Ying Liang**[1]

Received: 12 November 2014 / Revised: 7 May 2015 / Published online: 30 May 2015

Abstract For the partially linear varying-coefficient model when the parametric covariates are measured with additive errors, the estimator of the error variance is defined based on residuals of the model. At the same time, we construct Jackknife estimator as well as Jackknife empirical likelihood statistic of the error variance. Under both the response variables and their associated covariates form a stationary α-mixing sequence, we prove that the proposed estimators and Jackknife empirical likelihood statistic are asymptotic normality and asymptotic χ^2 distribution, respectively. Numerical simulations are carried out to assess the performance of the proposed method.

Keywords Asymptotic normality · Error variance · Jackknife empirical likelihood · Varying-coefficient errors-in-variables model · α-Mixing

Mathematics Subject Classification 62E20 · 62J10

1 Introduction

Consider the following partially linear varying-coefficient errors-in-variables (EV) model

$$\begin{cases} Y_i = X_i^\tau \beta + W_i^\tau a(T_i) + \epsilon_i, \\ \xi_i = X_i + e_i, \end{cases} \quad i = 1, 2, \ldots, n, \tag{1.1}$$

✉ Han-Ying Liang
hyliang83@yahoo.com

1 Department of Mathematics, Tongji University, Shanghai 200092, People's Republic of China

where Y_i are the scalar response variables and $(X_i^\tau, W_i^\tau, T_i)$ are covariates, $a(\cdot) = (a_1(\cdot), \cdots, a_q(\cdot))^\tau$ is a q-dimensional vector of unknown coefficient functions, $\beta = (\beta_1, \cdots, \beta_p)^\tau$ is a p-dimensional vector of unknown parameters, ϵ_i are random errors. Because of the curse of dimensionality, we assume that T_i is univariate; e_i are independent and identically distributed (i.i.d.) with mean zero and covariate matrix Σ_e, and are independent of (Y_i, X_i, W_i, T_i). In order to identify the model, Σ_e is assumed to be known. When Σ_e is unknown, one can employ the approaches proposed by Liang et al. (1999) to estimate it.

When X_i are observed exactly, the model (1.1) boils down to the partially linear varying-coefficient model, which has been studied by many authors, for example, Fan and Huang (2005) proposed a profile least square method to estimate the unknown parameter and studied the asymptotic normality of the estimator. Besides, based on the estimator, they introduced the profile empirical likelihood ratio test and showed the test statistic asymptotically χ^2 distributed under the null hypothesis. In addition, Ahmad et al. (2005), You and Zhou (2006), Huang and Zhang (2009), Wang et al. (2011), Bravo (2014) extensively explored partially linear varying-coefficient models; Zhou et al. (2010), Wei et al. (2012), Singh et al. (2014) for similar research related to EV models.

For the model (1.1), You and Chen (2006) studied the case where the covariates were observed with measurement errors and proposed estimators for the parametric and nonparametric component respectively. When the covariates in nonparametric part are measured with errors, Feng and Xue (2014) investigated the profile least square estimators and conducted a linear hypothesis test for the parametric part.

It is worth pointing out that the works mentioned above all assume that variables or errors are independent. However, the independence assumption is inadequate in some applications, especially in the field of economics and financial analysis, where the data often exhibit dependence to some extent. Therefore, the dependence data have drawn considerable interests of statisticians. One case of them is serially correlated errors, such as AR(1) errors, MA(∞) errors, negatively associated errors, martingale difference errors, etc. See, for example, the work of You et al. (2005), Liang et al. (2006), Liang and Jing (2009), You and Chen (2007), Fan et al. (2013), Fan et al. (2013) and Miao et al. (2013).

As we know, the empirical likelihood (EL) introduced by Owen (1988, 1990) is an effective method for constructing confidence regions which enjoys numerous nice properties over normal approximation-based methods and the bootstrap [see Hall (1992), Hall and La Scala (1990), Zi et al. (2012)]. The EL related to model (1.1) or partially linear varying-coefficient model has been studied by some authors, for example, You and Zhou (2006), Huang and Zhang (2009), Wang et al. (2011), and Fan et al. (2012) for the partially time-varying coefficient (in this case $T_i = i/n$) errors-in-variables model. It can be seen that the EL in these papers is based on linear functional of the studied parametric or nonparametric parts in the models. However, when nonlinear functionals are involved, such as U-statistics and variance of random sample, an application of the EL method will be computationally difficult and the Wilks theorem does not hold in general, i.e., the asymptotic distribution of the EL ratio is not a chi-squared distribution. Fortunately, in the study of the EL on one and two-sample U-statistics, Jing et al. (2009) proposed a new approach called jackknife empirical likelihood (JEL), which can handle the situation where nonlinear statistics are involved. At the same time, another attractive feature of the JEL is that the new method is simple to use. Thanks to the advantages, the JEL method has been applied recent years. See, for example, Gong et al. (2010), Peng (2012), Peng et al. (2012) and Feng and Peng (2012).

In the sequel, we assume that $\{(X_i, W_i, T_i, \epsilon_i), i \geq 1\}$ is a sequence of stationary α-mixing random variables with $E(\epsilon_i|X_i, W_i, T_i) = 0\ a.s.$ and $E(\epsilon_i^2|X_i, W_i, T_i) = \sigma^2\ a.s.$ from the model (1.1). Recall that a sequence $\{\zeta_k, k \geq 1\}$ is said to be α-mixing if the α-mixing coefficient

$$\alpha(n) :\overset{\text{def}}{=} \sup_{k\geq 1} \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}_{n+k}^{\infty}, B \in \mathcal{F}_1^k\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m = \sigma\{\zeta_l, \zeta_{l+1}, \cdots, \zeta_m\}$ denotes the σ-algebra generated by $\zeta_l, \zeta_{l+1}, \ldots, \zeta_m$ with $l \leq m$. As we know, among the most frequently used mixing conditions, the α-mixing is the weakest and many time series present α-mixing property. For a more detailed and general review, we refer to Doukhan (1994) and Lin and Lu (1996).

In this paper, we focus on estimating the error variance σ^2, and investigate asymptotic normality of estimator for the error variance. It is well known that the error of a regression model impacts its performance, and the study for the error variance could help researchers to improve the accuracy of the model. So it is necessary to investigate large sample properties of the estimators of the error variance. Up to now, only a few researchers have discussed the asymptotic normality of the estimator for the error variance. Among of them, we refer to You and Chen (2006), Liang and Jing (2009), Zhang and Liang (2012) and Fan et al. (2013), Fan et al. (2013). At the same time, we construct Jackknife estimator as well as JEL statistic of σ^2, and prove that they are asymptotic normality and asymptotic χ^2 distribution, respectively. Based on the JEL statistic of σ^2, we can construct its confidence interval which plays a crucial role in quantifying estimation uncertainty. With the study for error variance, we can get more comprehensive understanding of statistical models. Hence, the statistical inference can be improved. These results are new, even for independent data.

We organize the paper as follows. In Sect. 2, we give the methodologies and show how to build the estimators. Main results are listed in Sect. 3. Section 4 presents a simulation study to verify the idea and demonstrate the advantages of jackknife method. Proofs of Main Results are put in Sect. 5. Some preliminary lemmas, which are used in the proofs of the main results, are collected in Appendix.

2 Estimators

2.1 Profile least squares estimation

The local linear regression technique is applied to estimate the coefficient functions $\{a_j(\cdot), j = 1, 2, \cdots, q\}$ in (1.1). For t in a small neighborhood of t_0, one can approximate $a(t)$ locally by a linear function $a_j(t) \approx a_j(t_0) + a_j'(t_0)(t - t_0) \equiv a_j^* + b_j^*(t - t_0)$,

$j = 1, 2, \cdots, q$, where $a_j'(t) = \partial a_j(t)/\partial t$. This leads to the following weighted local least-squares problem if β is known: find (a^*, b^*) so as to minimize

$$\sum_{i=1}^{n} \Big[Y_i - X_i^\tau \beta - \Big(W_i^\tau, \ \ \frac{T_i - t}{h} W_i^\tau \Big) \Big(\begin{matrix} a^* \\ hb^* \end{matrix} \Big) \Big]^2 K_h(T_i - t), \tag{2.1}$$

where $a^* = (a_1^*, a_2^*, \cdots, a_q^*)^\tau$, $b^* = (b_1^*, b_2^*, \cdots, b_q^*)^\tau$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and $0 < h := h_n \to 0$ is a bandwidth.

For the sake of descriptive convenience, we denote $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)^\tau$, $\mathbf{X} = (X_1, X_2, \cdots, X_n)^\tau$, $\mathbf{W} = (W_1, W_2, \cdots, W_n)^\tau$, $\omega_t = diag(K_h(T_1 - t), K_h(T_2 - t), \cdots, K_h(T_n - t))$, and

$$M = \begin{pmatrix} W_1^\tau a(T_1) \\ \vdots \\ W_n^\tau a(T_n) \end{pmatrix}, \ \ D_t = \begin{pmatrix} W_1^\tau & \frac{T_1 - t}{h} W_1^\tau \\ \vdots & \vdots \\ W_n^\tau & \frac{T_n - t}{h} W_n^\tau \end{pmatrix}.$$

Then the minimizer in (2.1) is found to be $\Big(\begin{matrix} \hat{a}^* \\ h\hat{b}^* \end{matrix} \Big) = \{D_t^\tau \omega_t D_t\}^{-1} D_t^\tau \omega_t (\mathbf{Y} - \mathbf{X}\beta)$. Therefore, when β is known, we obtain the estimator of $\alpha(t)$ by

$$\tilde{a}(t, \beta) = \Big(I_q, \ \ 0_q \Big) \{D_t^\tau \omega_t D_t\}^{-1} D_t^\tau \omega_t (\mathbf{Y} - \mathbf{X}\beta). \tag{2.2}$$

Let $S_i = \Big(W_i^\tau \ \ 0 \Big) \{D_{T_i}^\tau \omega_{T_i} D_{T_i}\}^{-1} D_{T_i}^\tau \omega_{T_i}$, $\tilde{Y}_i = Y_i - S_i \mathbf{Y}$ and $\tilde{X}_i^\tau = X_i^\tau - S_i \mathbf{X}$. Substituting (2.2) into the original varying-coefficient model, and applying the least square method, one can obtain the estimator of parametric component β, $\tilde{\beta} = (\sum_{i=1}^n \tilde{X}_i \tilde{X}_i^\tau)^{-1} \sum_{i=1}^n \tilde{X}_i \tilde{Y}_i$. However, since X_i cannot be observed directly and we have $\xi_i = X_i + e_i$ instead, we can write (2.1) as

$$\sum_{i=1}^{n} \Big[Y_i - \xi_i^\tau \beta - \Big(W_i^\tau, \ \ \frac{T_i - t}{h} W_i^\tau \Big) \Big(\begin{matrix} a^* \\ hb^* \end{matrix} \Big) \Big]^2 K_h(T_i - t) - n\beta^\tau \Sigma_e \beta.$$

Similarly, one can obtain the following modified profile least squares estimator of β

$$\hat{\beta}_n = \Big(\sum_{i=1}^{n} \tilde{\xi}_i \tilde{\xi}_i^\tau - n\Sigma_e \Big)^{-1} \sum_{i=1}^{n} \tilde{\xi}_i \tilde{Y}_i,$$

and the estimators of $a(\cdot)$ and σ^2, respectively

$$\hat{a}_n(t) = \Big(I_q, \ \ 0_q \Big) \{D_t^\tau \omega_t D_t\}^{-1} D_t^\tau \omega_t (\mathbf{Y} - \xi \hat{\beta}_n),$$

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} [Y_i - \xi_i^\tau \hat{\beta}_n - W_i^\tau \hat{a}_n(T_i)]^2 - \hat{\beta}_n^\tau \Sigma_e \hat{\beta}_n.$$

2.2 Jackknife method

Since the estimators we have constructed are based on samples $(\tilde{\xi}_i, \tilde{Y}_i)_{i=1}^n$, they are regarded as the pseudo observations. Let $\hat{\beta}_{n,-i}$ be the estimator of β when the ith observation is deleted,

$$\hat{\beta}_{n,-i} = \Big[\sum_{j\neq i}^{n} \tilde{\xi}_j \tilde{\xi}_j^\tau - (n-1)\Sigma_e\Big]^{-1} \sum_{j\neq i}^{n} \tilde{\xi}_j \tilde{Y}_j.$$

Therefore the ith Jackknife pseudo sample is $J_i = n\hat{\beta}_n - (n-1)\hat{\beta}_{n,-i}$. Hence, we have the Jackknife estimator of β

$$\hat{\beta}_J = \frac{1}{n}\sum_{i=1}^{n} J_i = n\hat{\beta}_n - \frac{(n-1)}{n}\sum_{i=1}^{n}\hat{\beta}_{n,-i}.$$

From $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n)^2 - \hat{\beta}_n^\tau \Sigma_e \hat{\beta}_n$, similarly, let $\hat{\sigma}_{n,-i}^2$ be the estimator of σ^2 when the ith observation is deleted, $\hat{\sigma}_{n,-i}^2 = \frac{1}{n-1}\sum_{j\neq i}^{n}(\tilde{Y}_j - \tilde{\xi}_j^\tau \hat{\beta}_{n,-i})^2 - \hat{\beta}_{n,-i}^\tau \Sigma_e \hat{\beta}_{n,-i}$. Then we have the ith Jackknife pseudo sample $\sigma_{J_i}^2 = n\hat{\sigma}_n^2 - (n-1)\hat{\sigma}_{n,-i}^2$, and the Jackknife estimator of σ^2

$$\hat{\sigma}_J^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_{J_i}^2 = n\hat{\sigma}_n^2 - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\sigma}_{n,-i}^2.$$

Based on the Jackknife pseudo sample, one constructs the Jackknife empirical likelihood of σ^2

$$L(\sigma^2) := \sup\Big\{\prod_{i=1}^{n} np_i : p_1 > 0, p_2 > 0, \ldots, p_n > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \sigma_{J_i}^2 = \sigma^2\Big\}.$$

The solution to the above maximization is $\hat{p}_i = \frac{1}{n[1+\lambda(\sigma_{J_i}^2 - \sigma^2)]}$, $i = 1, 2, \ldots, n$, where λ satisfies $\frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_{J_i}^2 - \sigma^2}{1+\lambda(\sigma_{J_i}^2 - \sigma^2)} = 0$. Therefore, we have the log empirical likelihood ratio function of σ^2

$$l(\sigma^2) = 2\sum_{i=1}^{n}\log[1 + \lambda(\sigma_{J_i}^2 - \sigma^2)].$$

3 Main results

In order to formulate the main results, we need to impose the following basic assumptions.

(A1) The random variable T has bounded support Ω, and its density function $f(\cdot)$ is Lipschitz continuous and away from 0 on its support.
(A2) The $q \times q$ matrix $E(\mathbf{W}\mathbf{W}^\tau|T)$ is nonsingular for each $T \in \Omega$. $E(\mathbf{X}\mathbf{X}^\tau|T)$, $E(\mathbf{W}\mathbf{W}^\tau|T)$ and $E(\mathbf{X}\mathbf{W}^\tau|T)$ are all lipschitz continuous. Set $\Gamma(T_i) = E(W_i W_i^\tau|T_i)$, $\Phi(T_i) = E(X_i W_i^\tau|T_i)$, $i = 1, 2, \cdots, n$, the derivatives of order 2 of functions $\Gamma(\cdot)$ and $\Phi(\cdot)$ are bounded for each $T \in \Omega$. The $q \times q$ matrix $EX_1X_1^\tau - E\Phi^\tau(T_1)\Gamma^{-1}(T_1)\Phi(T_1)$ is positive definite.
(A3) There is a $\delta > 4$ such that $E(\|X_1\|^{2\delta}|T_1) < \infty\ a.s.$, $E(\|W_1\|^{2\delta}|T_1) < \infty\ a.s.$, $E\|\xi_1\|^{2\delta} < \infty\ a.s.$, $E[|\epsilon_1|^{2\delta}|X_1, W_1] < \infty\ a.s.$
(A4) $\{a_j(\cdot), j = 1, 2, \cdots, q\}$ have continuous second derivatives in $T \in \Omega$.
(A5) The function $K(\cdot)$ is a symmetric probability density function with bounded compact support which is Lipschitz continuous as well, and the bandwidth h satisfies $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$.
(A6) The α-mixing coefficient $\alpha(n)$ satisfies that $\alpha(n) = O(n^{-\lambda})$ for some $\lambda > \max\{\frac{7\delta+4}{\delta-4}, \frac{9\delta+4}{\delta+4}\}$ with the same δ as in (A3).

Remark 3.1 (a) Assumptions (A1)–(A6) are quite mild and commonly used in literature. Particularly, (A1)–(A2) and (A4)–(A5) are employed in Fan and Huang (2005), Feng and Xue (2014).
(b) Assumptions (A3) implies $E\|X_1\|^{2\delta} < \infty$ and $E\|W_1\|^{2\delta} < \infty$.
(c) Assumption (A6) indicates relatively low mixing speed. In fact, when the α-mixing coefficient decays exponentially, i.e. $\alpha(n) = O(\rho^n)$, $0 < \rho < 1$, one can verify easily that (A6) is satisfied.

Theorem 3.1 (i) *Suppose assumptions* (A1)–(A6) *are satisfied, then* $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{D}} N(0, \Pi)$, *where* $\Pi = \lim_{n\to\infty} Var\{\frac{1}{\sqrt{n}}\sum_{i=1}^n(\epsilon_i - e_i^\tau\beta)^2\}$. *Further,* $\hat{\Pi}$ *is a plug-in estimator of* Π, *where* $\hat{\Pi} = \frac{1}{n}\{\sum_{i=1}^n[(\tilde{Y}_i - \tilde{\xi}_i^\tau\hat{\beta}_n)^2 - \hat{\beta}_n^\tau\Sigma_e\hat{\beta}_n - \hat{\sigma}_n^2]\}^2$.
(ii) *Suppose assumptions* (A1)–(A6) *are satisfied, then* $\sqrt{n}(\hat{\sigma}_J^2 - \sigma^2) = \sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) + o_p(1)$. *Furthermore, with (i) we have* $\sqrt{n}(\hat{\sigma}_J^2 - \sigma^2) \xrightarrow{\mathcal{D}} N(0, \Pi)$.

Theorem 3.2 *Suppose assumptions* (A1)–(A6) *are satisfied, then* $\frac{\Sigma_4}{\Pi}l(\sigma^2) \xrightarrow{\mathcal{D}} \chi_1^2$, *where* $\Sigma_4 = E(\epsilon_1 - e_1^\tau\beta)^4 - (\sigma^2 + \beta^\tau\Sigma_e\beta)^2 > 0$. *Moreover,* $\hat{\Sigma}_4$ *is a plug-in estimator of* Σ_4, *where* $\hat{\Sigma}_4 = \frac{1}{n}\sum_{i=1}^n\{(\tilde{Y}_i - \tilde{\xi}_i^\tau\hat{\beta}_n)^4 - (\hat{\beta}_n^\tau\Sigma_e\hat{\beta}_n + \hat{\sigma}_n^2)^2\}$.

Remark 3.2 (a) Under the conditions of Theorem 3.2, if $\{\epsilon_i\}$ is a sequence of independent random variables, then one can verify $\Pi = \Sigma_4$ and $l(\sigma^2) \xrightarrow{\mathcal{D}} \chi_1^2$. In this case, the jackknife empirical likelihood method does not relate to estimation for the asymptotic variance Σ_4 of the jackknife pseudo samples. However, when $\{\epsilon_i\}$ is a sequence of dependent random variables, we cannot ignore the covariance between $(\epsilon_i - e_i^\tau\beta)^2$ and $(\epsilon_j - e_j^\tau\beta)^2)$ for $i \neq j$, which leads to $\Pi \neq \Sigma_4$. Thus, to construct an approximate confidence interval of σ^2, we need to estimate Π and Σ_4.
(b) From Theorem 3.2, it is easy to construct an approximate confidence region with level $1 - \tau$ for σ^2 as $I(\tau) = \{\sigma^2 : \frac{\hat{\Sigma}_4}{\hat{\Pi}}l(\sigma^2) \leq c_\tau\}$, where c_τ is chosen to satisfy $P(\chi_1^2 \leq c_\tau) = 1 - \tau$.

4 Simulation

In this section, we conduct numerical simulation to investigate the finite sample behavior of the profile least square estimator $\hat{\sigma}_n^2$ and the jackknife estimator $\hat{\sigma}_J^2$ in terms of sample means, bias, mean square error (MSE). Besides, we study the performance of proposed jackknife empirical likelihood method for constructing confidence intervals for σ^2 and compare it with normal approximation method in terms of coverage probability and average interval length.

Consider the following partially linear varying-coefficient EV model:

$$\begin{cases} Y_i = X_{1i}\beta_1 + X_{2i}\beta_2 + W_{1i}a_1(T_i) + W_{2i}a_2(T_i) + \epsilon_i, \\ \xi_i = X_i + e_i, \end{cases} \quad i = 1, 2, \ldots, n,$$

where $\beta_1 = 1$, $\beta_2 = 2$, $a_1(T) = sin(6\pi T)$, $a_2 = sin(2\pi T)$. The measurement error $e_i \sim N(0, \Sigma_e)$, where $\Sigma_e = 0.3^2 I_2$ and I_2 is the 2×2 identity matrix. $X_i, W_i, T_i, \epsilon_i$ are generated from AR(1) model as follows:

$X_{i,j} = \rho X_{i,j-1} + u_{i,j}, i = 1, 2$ with $u_{i,j}$ are i.i.d. $N(0, 1)$,
$W_{i,j} = \rho^2 W_{i,j-1} + w_{i,j}, i = 1, 2$ with $w_{i,j}$ are i.i.d. $N(0, 1)$,
$T_j = \sqrt{\rho} T_{j-1} + t_j$, t_j are i.i.d. $N(0, 0.1^2)$,
$\epsilon_j = \rho\epsilon_{j-1} + \eta_j$, η_j are i.i.d. $N(0, 0.5)$.

It is easy to verify that $\{X_i, W_i, T_i, \epsilon_i\}$ is a sequence of stationary and α-mixing random variables (see Doukhan (1994)) with $0 < \rho < 1$. When $\rho = 0$, $\{(X_i, W_i, T_i, \epsilon_i),\ \ i = 1, 2, \ldots, n\}$ are i.i.d. random variables. In order to investigate the influence of dependence on the estimators, we take the samples with ρ=0, 0.2, 0.5, 0.8, respectively. In fact, since the data generated from AR(1) model, one can easily find that the true value of $\sigma^2 = 0.5/(1 - \rho^2)$, which means that when the coefficient ρ changes, σ^2 changes as well.

The following simulation is based 1000 replications. For the proposed estimators, we employ the Epanechnikov kernel function $K(u) = 15/16(1-u^2)^2 I(|u| \leq 1)$, and the bandwidth h is selected by minimizing the MSE in a grid search.

Taking sample sizes $n = 50, 100, 200, 500$, we calculate bias and MSE of $\hat{\sigma}_n^2$ and $\hat{\sigma}_J^2$, respectively, to evaluate the two estimators' performance. According to Table 1, basically, the jackknife estimator performs better than the profile least square estimator. Both Bias($\hat{\sigma}_J^2$) and MSE($\hat{\sigma}_J^2$) are smaller than those of $\hat{\sigma}_n^2$. Besides, both estimators get more accurate when n increases. The gap between $MSE(\hat{\sigma}_n^2)$ and $MSE(\hat{\sigma}_J^2)$ becomes narrow as n increasing. In other words, the jackknife estimator can significantly improve the estimation accuracy when sample size is small. In addition, as the dependence of observations increases (i.e., ρ increases), which leads to larger σ^2, the accuracy of estimation slightly decreases when observations present relatively strong dependence. Specifically, the MSE for both estimators become larger as σ^2 rise.

Coverage probabilities and average interval lengths are reported in Table 2, showing that the jackknife empirical likelihood method is much more accurate than the normal approximation method in all scenarios in terms of coverage probabilities. Since it is obvious that the coverage probabilities for JEL are closer to the level than normal

Table 1 Sample means, biases and mean square errors for the estimator $\hat{\sigma}_n^2$ and $\hat{\sigma}_J^2$

ρ	σ^2	n	$\hat{\sigma}_n^2$	$\hat{\sigma}_J^2$	Bias($\hat{\sigma}_n^2$)	Bias($\hat{\sigma}_J^2$)	MSE($\hat{\sigma}_n^2$)$\times 10^3$	MSE($\hat{\sigma}_J^2$)$\times 10^3$
0	0.5	50	0.4362	0.4796	−0.0638	−0.0204	41.2140	40.6612
		100	0.4745	0.4886	−0.0255	−0.0114	19.3856	19.3127
		200	0.4854	0.4961	−0.0146	−0.0039	8.7411	8.7065
		500	0.4993	0.4998	−0.0007	−0.0002	3.8668	3.8435
0.2	0.5208	50	0.4699	0.4877	−0.0509	−0.0331	42.3700	41.2085
		100	0.4967	0.5027	−0.0241	−0.0181	20.1317	20.0729
		200	0.5159	0.5188	−0.0049	−0.0021	10.1983	10.0918
		500	0.5172	0.5195	−0.0036	−0.0013	4.2966	4.2889
0.5	0.6667	50	0.5946	0.6047	−0.0721	−0.0620	68.5674	67.2362
		100	0.6376	0.6446	−0.0291	−0.0221	32.0382	31.9127
		200	0.6568	0.6598	−0.0099	−0.0069	12.8815	12.8623
		500	0.6641	0.6648	−0.0026	−0.0019	7.3555	7.2993
0.8	1.3889	50	1.0907	1.1577	−0.2982	−0.2312	294.0034	276.1109
		100	1.2874	1.3022	−0.1015	−0.0867	175.6427	175.4381
		200	1.3274	1.3458	−0.0615	−0.0431	89.6432	89.4067
		500	1.3780	1.3782	−0.0109	−0.0107	38.3630	38.1840

Table 2 Coverage probabilities for the jackknife empirical likelihood (CP_J) and the normal approximation method based on σ_n^2 (CP_N) with confidence level 0.90, 0.95, respectively, and their corresponding average interval lengths AIL_J and AIL_N

ρ	n	Level 90 %				Level 95 %			
		CP_J	CP_N	AIL_J	AIL_N	CP_J	CP_N	AIL_J	AIL_N
0	100	0.887	0.873	0.4362	0.4425	0.935	0.927	0.5226	0.5258
	200	0.892	0.883	0.3116	0.3130	0.937	0.933	0.3724	0.3733
	500	0.897	0.889	0.1942	0.1981	0.942	0.938	0.2354	0.2374
0.2	100	0.855	0.842	0.4459	0.4478	0.915	0.902	0.5333	0.5341
	200	0.884	0.872	0.3188	0.3203	0.936	0.931	0.3806	0.3827
	500	0.889	0.881	0.2033	0.2078	0.939	0.934	0.2431	0.2444
0.5	100	0.835	0.812	0.5125	0.5129	0.908	0.876	0.6140	0.6111
	200	0.860	0.858	0.3669	0.3679	0.926	0.921	0.4401	0.4381
	500	0.893	0.891	0.2384	0.2392	0.930	0.927	0.2772	0.2764
0.8	100	0.789	0.746	0.8524	0.8685	0.863	0.846	1.0382	1.0405
	200	0.793	0.749	0.6015	0.6034	0.882	0.853	0.7203	0.7247
	500	0.834	0.792	0.3822	0.3798	0.892	0.861	0.4544	0.4531

approximation method (NAM). In most cases, the average interval lengths based on JEL are smaller than NAM. More precisely, as n increases, the coverage probabilities for both JEL method and NAM become closer to the level, the confidence intervals for both methods becomes narrow. When $\rho = 0$ i.e. independent cases, JEL performs much better than NAM with higher coverage probabilities and shorter confidence intervals. When dependence increases, the coverage probabilities slightly fall down, due to the fact that stronger dependence leads to bigger variance σ^2.

5 Proofs of main results

Throughout this paper, let C, C_1, C_2 denote finite positive constants, whose values may change in different scenarios. Let $\mu_i = \int u^i K(u)du$, and $c_n = \{\log(n)/(nh)\}^{1/2} + h^2$. From (A5), one can easily verify that $c_n = o(n^{-1/4})$. Set $\epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^\tau$, $\mathbf{1}_n = (1, 1, \cdots, 1)^\tau$.

Proof of Theorem 3.1 (i) From Lemma 6.3, it follows that $\frac{1}{\sqrt{n}} \sum_{i=1}^n [(\epsilon_i - e_i^\tau \beta)^2 - (\sigma^2 + \beta^\tau \Sigma_e \beta)] \stackrel{\mathcal{D}}{\to} N(0, \Pi)$, where $\Pi = \lim_{n\to\infty} Var\{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\epsilon_i - e_i^\tau \beta)^2\}$. Therefore, to prove Theorem 3.1 (i), it is sufficient to show that

$$\hat{\sigma}_n^2 - \sigma^2 = \frac{1}{n}\sum_{i=1}^n [(\epsilon_i - e_i^\tau \beta)^2 - (\sigma^2 + \beta^\tau \Sigma_e \beta)] + o_p(1).$$

From $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n)^2 - \hat{\beta}_n^\tau \Sigma_e \hat{\beta}_n$, one can write

$$\begin{aligned}\hat{\sigma}_n^2 - \sigma^2 &= \Big[\frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i^\tau \hat{\beta}_n)^2 - \sigma^2\Big] + \Big[\frac{1}{n}\sum_{i=1}^n \hat{\beta}_n^\tau \tilde{e}_i \tilde{e}_i^\tau \hat{\beta}_n - \hat{\beta}_n^\tau \Sigma_e \hat{\beta}_n\Big] \\ &\quad - \Big[\frac{2}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i^\tau \hat{\beta}_n)\hat{\beta}_n^\tau \tilde{e}_i\Big] \\ &:= A_1 + A_2 - A_3. \end{aligned} \tag{5.1}$$

First, we prove that

$$A_1 = \frac{1}{n}\sum_{i=1}^n (\epsilon_i^2 - \sigma^2) + o_p\Big(\frac{1}{\sqrt{n}}\Big), \tag{5.2}$$

$$A_2 = \frac{1}{n}\sum_{i=1}^n \beta^\tau (e_i e_i^\tau - \Sigma_e)\beta + o_p\Big(\frac{1}{\sqrt{n}}\Big), \tag{5.3}$$

$$A_3 = \frac{2}{n}\sum_{i=1}^n \epsilon_i e_i^\tau \beta + o_p\Big(\frac{1}{\sqrt{n}}\Big). \tag{5.4}$$

From the definition of $\tilde{Y}_i$ and (1.1), one can write

$$A_1 = \frac{1}{n}\sum_{i=1}^n (\epsilon_i^2 - \sigma^2) + \frac{1}{n}\sum_{i=1}^n [\tilde{X}_i^\tau (\beta - \hat{\beta}_n)]^2 + \frac{1}{n}\sum_{i=1}^n \tilde{M}_i^2 + \frac{1}{n}\sum_{i=1}^n (S_i \epsilon)^2$$

$$+\frac{2}{n}\sum_{i=1}^{n}[\tilde{X}_i^{\tau}(\beta-\hat{\beta}_n)]\tilde{M}_i+\frac{2}{n}\sum_{i=1}^{n}[\tilde{X}_i^{\tau}(\beta-\hat{\beta}_n)]\tilde{\epsilon}_i+\frac{2}{n}\sum_{i=1}^{n}\tilde{M}_i\tilde{\epsilon}_i-\frac{2}{n}\sum_{i=1}^{n}\epsilon_i S_i\epsilon$$

$$:=\frac{1}{n}\sum_{i=1}^{n}(\epsilon_i^2-\sigma^2)+\sum_{j=1}^{7}A_{1j}. \tag{5.5}$$

Note that from the proof of Lemma 3 in Owen (1990) and (A3), we have $\max_{1\le i\le n}\|X_i\|=o(n^{1/2\delta})\ \ a.s.$ and $\max_{1\le i\le n}\|W_i\|=o(n^{1/2\delta})\ \ a.s.$

Furthermore, from Lemma 6.6 and (A2), we have

$$\begin{aligned}\max_{1\le i\le n}\|\tilde{X}_i\| &\le \max_{1\le i\le n}\|X_i\|+\max_{1\le i\le n}\|W_i^{\tau}\Gamma^{-1}(T_i)\Phi(T_i)\|\{1+O_p(c_n)\}\\ &\le O_p(n^{1/2\delta})+C\max_{1\le i\le n}\|W_i^{\tau}\|\{1+O_p(c_n)\}=O_p(n^{1/2\delta}).\end{aligned}$$

Lemma 6.9 (i) gives $\|\hat{\beta}_n-\beta\|=O_p(n^{-1/2})$, therefore

$$A_{11}=\frac{1}{n}\sum_{i=1}^{n}[\tilde{X}_i^{\tau}(\beta-\hat{\beta}_n)]^2\le\max_{1\le i\le n}\|\tilde{X}_i\|^2\|\beta-\hat{\beta}_n\|^2=O_p(n^{1/\delta-1})=o_p(n^{-1/2}). \tag{5.6}$$

From (A1)–(A4), one can easily obtain that $P\big(\frac{1}{n}\sum_{i=1}^{n}(W_i^{\tau}a(T_i))^2>\eta\big)\le\frac{E[a^{\tau}(T_1)\Gamma(T_1)a(T_1)]}{\eta}<\frac{C}{\eta}$, which implies $\frac{1}{n}\sum_{i=1}^{n}(W_i^{\tau}a(T_i))^2=O_p(1)$. Together with (6.9) and (A5) we have

$$A_{12}=\frac{1}{n}\sum_{i=1}^{n}(W_i^{\tau}a(T_i))^2O_p(c_n^2)=O_p(c_n^2)=o_p(n^{-1/2}). \tag{5.7}$$

Note that $\frac{1}{n}\sum_{i=1}^{n}W_iW_i^{\tau}=O_p(1)$. Therefore, together with (6.14), we have

$$A_{13}=\frac{1}{n}\sum_{i=1}^{n}(S_i\epsilon)^2=\frac{1}{n}\sum_{i=1}^{n}W_i^{\tau}W_iO_p\Big(\frac{\log n}{nh}\Big)=O_p\Big(\frac{\log n}{nh}\Big)=o_p(n^{-1/2}). \tag{5.8}$$

From (6.9), (A3) and (A4), we have $\max_{1\le i\le n}|\tilde{M}_i|=\max_{1\le i\le n}|W_i^{\tau}a(T_i)|O_p(c_n)=O_p(n^{1/2\delta})O_p(c_n)$. Similar to the proof of (5.6), one can obtain that

$$|A_{14}|\le 2(\max_{1\le i\le n}\|\tilde{X}_i^{\tau}\|\|\beta-\hat{\beta}_n\|\max_{1\le i\le n}|\tilde{M}_i|)=O_p(n^{1/\delta-1/2}c_n)=o_p(n^{-1/2}). \tag{5.9}$$

As to A_{15}, by (6.6), (6.14), Lemma 6.10, (A1), (A2) and (A5), we have

$$|A_{15}|=\left|\frac{2}{n}\sum_{i=1}\tilde{X}_i^{\tau}(\beta-\hat{\beta}_n)\epsilon_i-\frac{2}{n}\sum_{i=1}^{n}\tilde{X}_i^{\tau}(\beta-\hat{\beta}_n)W_i^{\tau}O_p\Big(\sqrt{\frac{\log n}{nh}}\Big)\right|$$

$$\leq \left\| \frac{2}{n} \sum_{i=1} X_i^\tau \epsilon_i \right\| \left\| \beta - \hat{\beta}_n \right\| [1 + O_p(c_n)]$$
$$+ \left\| \frac{2}{n} \sum_{i=1} W_i^\tau \Gamma^{-1}(T_i) \Phi(T_i) \epsilon_i \right\| \left\| \beta - \hat{\beta}_n \right\| [1 + O_p(c_n)]$$
$$+ \max_{1 \leq i \leq n} \| \tilde{X}_i^\tau \| \max_{1 \leq i \leq n} \| W_i^\tau \| \| \beta - \hat{\beta}_n \| O_p\left(\sqrt{\frac{\log n}{nh}} \right)$$
$$= o(n^{-1/4}) O_p(n^{-1/2}) + O_p(n^{1/\delta}) O_p(n^{-1/2}) O_p\left(\sqrt{\frac{\log n}{nh}} \right) = o_p(n^{-1/2}). \tag{5.10}$$

From (A1), (A2), (A4), it is easy to verify that $|\frac{1}{n} a^\tau(T_i) W_i W_i^\tau \mathbf{1}| = O_p(1)$. Therefore, with Lemma 6.10, (6.9) and (6.14), we have

$$|A_{16}| = \left| \frac{2}{n} \sum_{i=1}^n a^\tau(T_i) W_i \tilde{\epsilon}_i \right| O_p(c_n) \leq \left| \frac{2}{n} \sum_{i=1}^n a^\tau(T_i) W_i \epsilon_i \right| O_p(c_n)$$
$$+ \left| \frac{2}{n} \sum_{i=1}^n a^\tau(T_i) W_i W_i^\tau \mathbf{1} \right| O_p(c_n) O_p\left(\sqrt{\frac{\log n}{nh}} \right) = o_p(n^{-1/2}). \tag{5.11}$$

From Lemma 6.10 and (6.14), it is directly derived that

$$|A_{17}| = \left| \frac{2}{n} \sum_{i=1}^n \epsilon_i W_i^\tau \mathbf{1} \right| O_p\left(\sqrt{\frac{\log n}{nh}} \right) = o(n^{-1/4}) O_p\left(\sqrt{\frac{\log n}{nh}} \right) = o_p(n^{-1/2}). \tag{5.12}$$

Hence, with (5.4)–(5.7),(5.8)–(5.12), we finish the proof of (5.2). Write

$$A_2 = \frac{1}{n} \sum_{i=1}^n \beta^\tau (e_i e_i^\tau - \Sigma_e) \beta \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_n - \beta)^\tau (e_i e_i^\tau - \Sigma_e)(\hat{\beta}_n - \beta)$$
$$+ \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_n - \beta)^\tau (e_i e_i^\tau - \Sigma_e) \beta + \frac{1}{n} \sum_{i=1}^n \beta^\tau (e_i e_i^\tau - \Sigma_e)(\hat{\beta}_n - \beta)$$
$$:= \frac{1}{n} \sum_{i=1}^n \beta^\tau (e_i e_i^\tau - \Sigma_e) \beta + A_{21} + A_{22} + A_{23}. \tag{5.13}$$

Note that $\frac{1}{n} \sum_{i=1}^n e_i e_i^\tau - \Sigma_e = o_p(1)$ from the strong law of large number for i.i.d. random variables and $\|\hat{\beta}_n - \beta\| = O_p(n^{-1/2})$. Then

$$|A_{21}| = \left| (\hat{\beta}_n - \beta)^\tau \left[\frac{1}{n} \sum_{i=1}^n e_i e_i^\tau - \Sigma_e \right] (\hat{\beta}_n - \beta) \right| = o_p(n^{-1}) = o_p(n^{-1/2}), \tag{5.14}$$

$$|A_{22}| = \left|(\hat{\beta}_n - \beta)^\tau \left[\frac{1}{n}\sum_{i=1}^n e_i e_i^\tau - \Sigma_e\right]\beta\right| = o_p(n^{-1/2}), \tag{5.15}$$

$$|A_{23}| = \left|\beta^\tau \left[\frac{1}{n}\sum_{i=1}^n e_i e_i^\tau - \Sigma_e\right](\hat{\beta}_n - \beta)\right| = o_p(n^{-1/2}). \tag{5.16}$$

Hence, by (5.13)–(5.16), we complete the proof of (5.3). Write

$$\begin{aligned} A_3 &= \frac{2}{n}\sum_{i=1}^n \epsilon_i e_i^\tau \beta - \frac{2}{n}\sum_{i=1}^2 S_i \epsilon e_i^\tau \beta + \frac{2}{n}\sum_{i=1}^n [\tilde{X}_i^\tau(\beta - \hat{\beta}_n) + \tilde{M}_i]\beta^\tau e_i \\ &\quad + \frac{2}{n}\sum_{i=1}^n [\tilde{X}_i^\tau(\beta - \hat{\beta}_n) + \tilde{M}_i + \tilde{\epsilon}_i](\hat{\beta}_n - \beta)^\tau e_i \\ &:= \frac{2}{n}\sum_{i=1}^n \epsilon_i e_i^\tau \beta + A_{31} + A_{32} + A_{33}. \end{aligned} \tag{5.17}$$

Applying Lemma 6.3, we have $\|\frac{1}{n}\sum_{i=1}^n W_i e_i^\tau\| = O_p(n^{-1/2})$. Then by (6.14), we have

$$|A_{31}| = \left|\frac{2}{n}\sum_{i=1}^n \mathbf{1}^\tau W_i e_i^\tau \beta\right| O_p\left(\sqrt{\frac{\log n}{nh}}\right) = O_p(n^{-1/2}) O_p\left(\sqrt{\frac{\log n}{nh}}\right) = o_p(n^{-1/2}). \tag{5.18}$$

Similarly, by (6.6) and (6.9), one can obtain that

$$\begin{aligned} |A_{32}| &= \left|(\beta - \hat{\beta}_n)^\tau \left[\frac{2}{n}\sum_{i=1}^n \tilde{X}_i e_i^\tau\right]\beta + \left[\frac{2}{n}\sum_{i=1}^n \tilde{M}_i e_i^\tau\right]\beta\right| \\ &\le \left|(\beta - \hat{\beta}_n)^\tau \left[\frac{2}{n}\sum_{i=1}^n X_i e_i^\tau\right]\beta\right|[1 + O_p(c_n)] \\ &\quad + \left|(\beta - \hat{\beta}_n)^\tau \left[\frac{2}{n}\sum_{i=1}^n \Phi(T_i)\Gamma^{-1}(T_i) W_i e_i^\tau\right]\beta\right|[1 + O_p(c_n)] \\ &\quad + \left|\left[\frac{2}{n}\sum_{i=1}^n a^\tau(T_i) W_i e_i^\tau\right]\beta\right| O_p(c_n) = o_p(n^{-1/2}), \end{aligned} \tag{5.19}$$

$$\begin{aligned} |A_{33}| &= \left|(\beta - \hat{\beta}_n)^\tau \left[\frac{2}{n}\sum_{i=1}^n \tilde{X}_i e_i^\tau\right](\beta - \hat{\beta}_n)\right| + \left|\left[\frac{2}{n}\sum_{i=1}^n \tilde{M}_i e_i^\tau\right](\hat{\beta}_n - \beta)\right| \\ &\quad + \left|\left[\frac{2}{n}\sum_{i=1}^n \tilde{\epsilon}_i e_i^\tau\right](\hat{\beta}_n - \beta)\right| \end{aligned}$$

$$
\begin{aligned}
&\le O_p(n^{-1})O_p(n^{-1/2}) + O_p(n^{-1})O_p(c_n) + \left|\frac{2}{n}\sum_{i=1}^{n}\epsilon_i e_i^\tau(\hat\beta_n - \beta)\right| \\
&\quad + \left|\frac{2}{n}\sum_{i=1}^{n} S_i \epsilon e_i^\tau(\hat\beta_n - \beta)\right| \\
&= O_p(n^{-3/2}) + O_p(n^{-1}c_n) + O_p(n^{-1}) + O_p(n^{-1})O_p\left(\sqrt{\frac{\log n}{nh}}\right) = o_p(n^{-1/2}).
\end{aligned} \tag{5.20}
$$

Combining (5.17)–(5.20), we prove (5.4). As a result, (5.1) can be written as

$$
\hat\sigma_n^2 - \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}[(\epsilon_i - e_i^\tau\beta)^2 - (\sigma^2 + \beta^\tau\Sigma_e\beta)] + o_p(n^{-1/2}).
$$

This completes the proof of Theorem 3.1 (i).

(ii) To prove $\sqrt{n}(\hat\sigma_J^2 - \sigma^2) = \sqrt{n}(\hat\sigma_n^2 - \sigma^2) + o_p(1)$, it is sufficient to prove that $\hat\sigma_J^2 = \hat\sigma_n^2 + o_p(n^{-1/2})$. According to the definition, we have $\hat\sigma_J^2 = \hat\sigma_n^2 + \frac{n-1}{n}\sum_{i=1}^n(\hat\sigma_n^2 - \hat\sigma_{n,-i}^2)$. Therefore, to obtain the desired result, we only need to prove

$$
\sqrt{n}\sum_{i=1}^{n}(\hat\sigma_n^2 - \hat\sigma_{n,-i}^2) = o_p(1). \tag{5.21}
$$

Note that $\sum_{i=1}^n[\tilde\xi_i(\tilde Y_i - \tilde\xi_i^\tau\hat\beta_n) + \Sigma_e\hat\beta_n] = 0$, with Lemma 6.4 we have

$$
\begin{aligned}
\sum_{i=1}^{n}(\hat\sigma_n^2 - \hat\sigma_{n,-i}^2) &= \frac{1}{n-1}\sum_{i=1}^{n}(\hat\beta_n - \hat\beta_{n,-i})^\tau\tilde\xi_i\tilde\xi_i^\tau(\hat\beta_n - \hat\beta_{n,-i}) \\
&\quad + \frac{2}{n-1}\sum_{i=1}^{n}[\tilde\xi_i^\tau(\tilde Y_i - \tilde\xi_i^\tau\hat\beta_n) + \hat\beta_n^\tau\Sigma_e](\hat\beta_n - \hat\beta_{n,-i}) \\
&\quad + \sum_{i=1}^{n}(\hat\beta_n - \hat\beta_{n,-i})^\tau\Sigma_e(\hat\beta_n - \hat\beta_{n,-i}) \\
&\quad - \frac{1}{n-1}\sum_{i=1}^{n}(\hat\beta_n - \hat\beta_{n,-i})^\tau\sum_{j=1}^{n}\tilde\xi_j\tilde\xi_j^\tau(\hat\beta_n - \hat\beta_{n,-i}) := \sum_{k=1}^{4}B_k.
\end{aligned}
$$

Therefore, to prove (5.21), it is sufficient to prove $B_k = o_p(n^{-1/2}),\quad k = 1, 2, 3, 4$.

From Lemmas 6.7 and 6.11, we have

$$
B_1 = \frac{1}{n-1}\sum_{i=1}^{n}[(\hat\beta_n - \hat\beta_{n,-i})^\tau\tilde\xi_i\tilde\xi_i^\tau(\hat\beta_n - \hat\beta_{n,-i})] = O_p(n^{-2}). \tag{5.22}
$$

Similarly, one can easily check that

$$B_3 = \sum_{i=1}^{n} (\hat{\beta}_n - \hat{\beta}_{n,-i})^\tau \Sigma_e (\hat{\beta}_n - \hat{\beta}_{n,-i}) O_p(n^{-1}), \tag{5.23}$$

$$B_4 = \sum_{i=1}^{n} (\hat{\beta}_n - \hat{\beta}_{n,-i})^\tau \frac{1}{n-1} \sum_{j=1}^{n} \tilde{\xi}_j \tilde{\xi}_j^\tau (\hat{\beta}_n - \hat{\beta}_{n,-i}) = O_p(n^{-1}). \tag{5.24}$$

Using Lemmas 6.11 and 6.5, we have

$$\begin{aligned} B_2^2 &= \frac{4}{(n-1)^2} \Big\{ \sum_{i=1}^{n} [\tilde{\xi}_i^\tau (\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \hat{\beta}_n^\tau \Sigma_e](\hat{\beta}_n - \hat{\beta}_{n,-i}) \Big\}^2 \\ &\le \frac{4n}{(n-1)^2} \sum_{i=1}^{n} (\tilde{\xi}_i^\tau (\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \hat{\beta}_n^\tau \Sigma_e)^2 O_p(n^{-2}) = O_p(n^{-2}). \end{aligned}$$

Therefore, one can obtain that

$$|B_2| = o_p(n^{-1/2}). \tag{5.25}$$

Hence, combining (5.22)–(5.25), we finish the proof of (5.21). □

Proof of Theorem 3.2. Define $g(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{\sigma_{J_i}^2 - \sigma^2}{1+\lambda(\sigma_{J_i}^2 - \sigma^2)}$. It is easy to check that

$$\begin{aligned} 0 = |g(\lambda)| &= \Big| \frac{1}{n} \sum_{i=1}^{n} (\sigma_{J_i}^2 - \sigma^2) - \frac{\lambda}{n} \sum_{i=1}^{n} \frac{(\sigma_{J_i}^2 - \sigma^2)^2}{1 + \lambda(\sigma_{J_i}^2 - \sigma^2)} \Big| \ge \frac{|\lambda| S_{\sigma^2}}{1 + |\lambda| R_n} \\ &\quad - \Big| \frac{1}{n} \sum_{i=1}^{n} \sigma_{J_i}^2 - \sigma^2 \Big|, \end{aligned}$$

where $S_{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (\sigma_{J_i}^2 - \sigma^2)^2$, $R_n = \max_{1 \le i \le n} |\sigma_{J_i}^2 - \sigma^2|$. Next we prove

$$R_n = \max_{1 \le i \le n} |\sigma_{J_i}^2 - \sigma^2| = o_p(\sqrt{n}), \tag{5.26}$$

$$S_{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (\sigma_{J_i}^2 - \sigma^2)^2 \xrightarrow{P} \Sigma_4. \tag{5.27}$$

Write

$$\begin{aligned} \hat{\sigma}_n^2 - \hat{\sigma}_{n,-i}^2 &= \frac{1}{n-1} [(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n)^2 - \hat{\sigma}_n^2 - \hat{\beta}_n^\tau \Sigma_e \hat{\beta}_n] \\ &\quad + \frac{1}{n-1} (\hat{\beta}_n - \hat{\beta}_{n,-i})^\tau \tilde{\xi}_i \tilde{\xi}_i^\tau (\hat{\beta}_n - \hat{\beta}_{n,-i}) \end{aligned}$$

$$+\frac{2}{n-1}[\tilde{\xi}_i^\tau(\tilde{Y}_i-\tilde{\xi}_i^\tau\hat{\beta}_n)+\hat{\beta}_n^\tau\Sigma_e](\hat{\beta}_n-\hat{\beta}_{n,-i})$$
$$+(\hat{\beta}_n-\hat{\beta}_{n,-i})^\tau\Sigma_e(\hat{\beta}_n-\hat{\beta}_{n,-i})$$
$$-\frac{1}{n-1}(\hat{\beta}_n-\hat{\beta}_{n,-i})^\tau\sum_{j=1}^n\tilde{\xi}_j\tilde{\xi}_j^\tau(\hat{\beta}_n-\hat{\beta}_{n,-i}):=\sum_{k=1}^5 b_{ki}.$$

Hence, to prove (5.26) we only need to prove $\max_{1\le i\le n}|b_{ki}|=o_p(n^{-1/2})$ for $k=1,2,3,4,5$.

Apparently, we have

$$\frac{(n-1)^2}{n}\sum_{i=1}^n b_{1i}^2=\frac{1}{n}\sum_{i=1}^n\Big[(\epsilon_i-e_i^\tau\beta)^4+(\tilde{\xi}_i^\tau(\beta-\hat{\beta}_n))^4+4(\epsilon_i-e_i^\tau\beta)^3\tilde{\xi}(\beta-\hat{\beta}_n)$$
$$+4(\epsilon_i-e_i^\tau\beta)(\tilde{\xi}(\beta-\hat{\beta}_n))^3+6(\epsilon_i-e_i^\tau\beta)^2(\tilde{\xi}(\beta-\hat{\beta}_n))^2\Big]$$
$$-(\hat{\sigma}_n^2+\hat{\beta}_n^\tau\Sigma_e\hat{\beta}_n)^2.$$

From (A3), we have

$$P\Big(n^{-3/2}\Big|\sum_{i=1}^n(\epsilon_i-e_i^\tau\beta)^3\tilde{\xi}_i\Big|>\eta\Big)\le\frac{1}{\eta}n^{-3/2}\sum_{i=1}^n E|(\epsilon_i-e_i^\tau\beta)^3\tilde{\xi}_i|\to 0,$$

which implies $\frac{4}{n}\sum_{i=1}^n(\epsilon_i-e_i^\tau\beta)^3\tilde{\xi}(\beta-\hat{\beta}_n)=o_p(1)$ from $\|\hat{\beta}_n-\beta\|=O_p(n^{-1/2})$ given by Lemma 6.9 (i). Similarly $\frac{4}{n}\sum_{i=1}^n(\epsilon_i-e_i^\tau\beta)(\tilde{\xi}(\beta-\hat{\beta}_n))^3=o_p(1)$, $\frac{6}{n}\sum_{i=1}^n(\epsilon_i-e_i^\tau\beta)^2(\tilde{\xi}(\beta-\hat{\beta}_n))^2=o_p(1)$ and $\frac{1}{n}\sum_{i=1}^n(\tilde{\xi}(\beta-\hat{\beta}_n))^4=o_p(1)$. Therefore, from Lemma 6.5, we have

$$\frac{(n-1)^2}{n}\sum_{i=1}^n b_{1i}^2\xrightarrow{\text{P}}E(\epsilon_1-e_1^\tau\beta)^4-(\sigma^2+\beta^\tau\Sigma_e\beta)^2=\Sigma_4. \tag{5.28}$$

From (5.28), one can derive that

$$\max_{1\le i\le n}|b_{1i}|=o_p(n^{-1/2}). \tag{5.29}$$

By the same approaches used in (5.22)-(5.25), one can easily check

$$\max_{1\le i\le n}|b_{ki}|=O_p(n^{-1}),\quad k=2,3,4,5. \tag{5.30}$$

Hence, together with (5.29) and (5.30), we have proved (5.26).

According to Theorem 3.1, one can write $S_{\sigma^2}=\frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2)^2-(\sigma^2)^2+o_p(1)$,

$$\frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2)^2=(\hat{\sigma}_n^2)^2+\frac{2(n-1)}{n}\hat{\sigma}_n^2\sum_{i=1}^n(\hat{\sigma}_n^2-\hat{\sigma}_{n,-i}^2)+\frac{(n-1)^2}{n}\sum_{i=1}^n(\hat{\sigma}_n^2-\hat{\sigma}_{n,-i}^2)^2.$$

Therefore, to prove (5.27), we need to investigate the convergency of $\frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2)^2$ first.

From (5.21), we have $\frac{2(n-1)}{n}\hat{\sigma}_n^2\sum_{i=1}^n(\hat{\sigma}_n^2 - \hat{\sigma}_{n,-i}^2) = o_p(n^{-1/2})$. Using the same techniques in proving (5.26), one can get $\frac{(n-1)^2}{n}\sum_{i=1}^n(\hat{\sigma}_n^2 - \hat{\sigma}_{n,-i}^2)^2 = \frac{(n-1)^2}{n}\sum_{i=1}^n b_{1i}^2 + o_p(1)$. Together with (5.28), we have

$$S_{\sigma^2} = \frac{(n-1)^2}{n}\sum_{i=1}^n b_{1i}^2 + o_p(1) \xrightarrow{\mathrm{P}} \Sigma_4,$$

which proves (5.27).

Applying Theorem 3.1, we have $|\frac{1}{n}\sum_{i=1}^n \sigma_{J_i}^2 - \sigma^2| = O_p(n^{-1/2})$. Together with (5.27), we have $\frac{|\lambda|}{1+|\lambda|R_n} = O_p(n^{-1/2})$. From (5.26), it follows that $|\lambda| = O_p(n^{-1/2})$. Let $\gamma_i = \lambda(\sigma_{J_i}^2 - \sigma^2)$, then still by (5.26), $\max_{1\le i\le n}|\gamma_i| = |\lambda|R_n = o_p(1)$. Note that

$$\begin{aligned}0 = g(\lambda) &= \frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2)\frac{1}{1+\gamma_i} = \frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2)(1 - \gamma_i + \frac{\gamma_i^2}{1+\gamma_i}) \\ &= \frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2) - \lambda S_{\sigma^2} + \frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2)\frac{\gamma_i^2}{1+\gamma_i}.\end{aligned}$$

By (5.26) and (5.27), it is easy to derive that $\frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2)\frac{\gamma_i^2}{1+\gamma_i} = \frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2)^2\lambda^2(\sigma_{J_i}^2 - \sigma^2)\frac{1}{1+\gamma_i} = o_p(n^{-1/2})$. Therefore

$$\lambda S_{\sigma^2} = \frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2) + o_p(n^{-1/2}).$$

Denote $\lambda = S_{\sigma^2}^{-1}\frac{1}{n}\sum_{i=1}^n(\sigma_{J_i}^2 - \sigma^2) + \phi_n$, where $|\phi_n| = o_p(n^{-1/2})$. Let $\eta_i = \sum_{k=3}^\infty \frac{(-1)^{k-1}}{k}\gamma_i^k$, then $\eta_i = O(\gamma_i^3)$, which implies $|\sum_{i=1}^n \eta_i| \le C\sum_{i=1}^n|\gamma_i|^3 = C\sum_{i=1}^n|\lambda^2(\hat{\sigma}_{J_i}^2 - \sigma^2)^2\gamma_i| \le Cn\lambda^2 S_{\sigma^2}\max_{1\le i\le n}|\gamma_i| = o_p(1)$. Hence

$$\begin{aligned}l(\sigma^2) &= 2\sum_{i=1}^n\gamma_i - \sum_{i=1}^n\gamma_i^2 + 2\sum_{i=1}^n\eta_i = 2\lambda n(\hat{\sigma}_J^2 - \sigma^2) - n\lambda^2 S_{\sigma^2} + 2\sum_{i=1}^n\eta_i \\ &= 2n(\hat{\sigma}_J^2 - \sigma^2)^2[S_{\sigma^2}^{-1}(\hat{\sigma}_J^2 - \sigma^2) + \phi_n] - nS_{\sigma^2}[S_{\sigma^2}^{-1}(\hat{\sigma}_J^2 - \sigma^2) + \phi_n]^2 + 2\sum_{i=1}^n\eta_i \\ &= nS_{\sigma^2}^{-1}(\hat{\sigma}_J^2 - \sigma^2)^2 - nS_{\sigma^2}\phi_n^2 + 2\sum_{i=1}^n\eta_i \\ &= nS_{\sigma^2}^{-1}(\hat{\sigma}_J^2 - \sigma^2)^2 + o_p(1).\end{aligned}$$

Finally, together with Theorem 3.1, we finish the proof of Theorem 3.2. □

Acknowledgements The authors would like to thank anonymous referees for their valuable comment and suggestions which lead to the improvement of the paper. This research was supported by the National Natural Science Foundation of China (11271286) and the Specialized Research Fund for the Doctor Program of Higher Education (20120072110007).

Appendix

In this section, we give some preliminary Lemmas, which have been used in Section 5. Let $\{X_i, i \geq 1\}$ be a stationary sequence of α-mixing random variables with the mixing coefficients $\{\alpha(k)\}$.

Lemma 6.1 (Liebscher (2001), Proposition 5.1) *Assume that* $EX_i = 0$ *and* $|X_i| \leq S < \infty$ *a.s.* $(i = 1, 2, \cdots, n)$. *Then for* $n, m \in \mathbb{N}$, $0 < m \leq n/2$ *and* $\epsilon > 0$, $P(|\sum_{i=1}^n X_i| > \epsilon) \leq 4\exp\{-\frac{\epsilon^2}{16}(nm^{-1}D_m + \frac{1}{3}\epsilon Sm)^{-1}\} + 32\frac{S}{\epsilon}n\alpha(m)$, *where* $D_m = \max_{1\leq j\leq 2m} Var(\sum_{i=1}^j X_i)$.

Lemma 6.2 (Yang (2007), Theorem 2.2)

(i) *Let* $r > 2$, $\delta > 0$, $EX_i = 0$ *and* $E|X_i|^{r+\delta} < \infty$. *Suppose that* $\lambda > r(r+\delta)/(2\delta)$ *and* $\alpha(n) = O(n^{-\lambda})$. *Then for any* $\epsilon > 0$, *there exists a positive constant* $C := C(\epsilon, r, \delta, \lambda)$ *such that* $E\max_{1\leq m\leq n}|\sum_{i=1}^m X_i|^r \leq C\{n^\epsilon \sum_{i=1}^n E|X_i|^r + (\sum_{i=1}^n \|X_i\|_{r+\delta}^2)^{r/2}\}$.

(ii) *If* $EX_i = 0$ *and* $E|X_i|^{2+\delta} < \infty$ *for some* $\delta > 0$, *then* $E(\sum_{i=1}^n X_i)^2 \leq \{1 + 16\sum_{l=1}^n \alpha^{\frac{\delta}{2+\delta}}(l)\}\sum_{i=1}^n \|X_i\|_{2+\delta}^2$.

Lemma 6.3 (Lin and Lu (1996), Theorem 3.2.1) *Suppose that* $EX_1 = 0$, $E|X_1|^{2+\delta} < \infty$ *for some* $\delta > 0$ *and* $\sum_{n=1}^\infty \alpha^{\delta/(2+\delta)}(n) < \infty$. *Then* $\sigma^2 := EX_1^2 + 2\sum_{j=2}^\infty EX_1X_j < \infty$ *and, if* $\sigma \neq 0$, $\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, 1)$.

Lemma 6.4 (Miller (1974), Lemma 2.1) *For a nonsingular matrix* A, *and vectors* U *and* V, *we have* $(A + UV^\tau)^{-1} = A^{-1} - \frac{(A^{-1}U)(V^\tau A^{-1})}{1+V^\tau A^{-1}U}$.

Lemma 6.5 (Shao (1993), Corollary 1) *Let* $EX_i = 0$ *and* $\sup_i E|X_i|^r < \infty$ *for some* $r > 1$. *Suppose that* $\alpha(n) = O(\log^{-\psi} n)$ *for some* $\psi > r/(r-1)$. *Then* $n^{-1}\sum_{i=1}^n X_i = o(1)$ *a.s.*

Lemma 6.6 *Suppose* (A1)–(A3), (A5) *and* (A6) *are satisfied, then*

$$\sup_{t\in\Omega}\left|\frac{1}{n}D_t^\tau\omega_t D_t - f(t)\Gamma(t)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\right| = O_p(c_n), \tag{6.1}$$

$$\sup_{t\in\Omega}\left|\frac{1}{n}D_t^\tau\omega_t X - f(t)\Phi(t)\otimes\begin{pmatrix}1\\ 0\end{pmatrix}\right| = O_p(c_n). \tag{6.2}$$

Proof We only prove (6.1) here, because (6.2) can be proved similarly. Write

$$D_t^\tau \omega_t D_t = \begin{pmatrix} W_1, & \dots, & W_n \\ \frac{T_1-t}{h} W_1, & \dots, & \frac{T_n-t}{h} W_n \end{pmatrix} \begin{pmatrix} K_h(T_1-t) & & \\ & \ddots & \\ & & K_h(T_n-t) \end{pmatrix} \begin{pmatrix} W_1^\tau & \frac{T_1-t}{h} W_1^\tau \\ \vdots & \vdots \\ W_n^\tau & \frac{T_n-t}{h} W_n^\tau \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{i=1}^n W_i W_i^\tau K_h(T_i - t) & \sum_{i=1}^n W_i W_i^\tau \frac{T_i-t}{h} K_h(T_i - t) \\ \sum_{i=1}^n W_i W_i^\tau \frac{T_i-t}{h} K_h(T_i - t) & \sum_{i=1}^n W_i W_i^\tau \left(\frac{T_i-t}{h}\right)^2 K_h(T_i - t) \end{pmatrix}. \tag{6.3}$$

Here, we only give the proof of

$$\sup_{t\in\Omega} \left| \frac{1}{n} \sum_{i=1}^n W_i W_i^\tau K_h(T_i - t) - f(t)\Gamma(t) \right| = O_p(c_n). \tag{6.4}$$

We divide Ω into subintervals $\{\Delta_l\}$ $(l = 1, 2, \cdots, l_n)$ with length $r_n = h\sqrt{\frac{\log n}{nh}}$, and the center of Δ_l is at t_l. Then the total number of the subintervals satisfies $l_n = O(r_n^{-1})$. Then

$$\begin{aligned} &\sup_{t\in\Omega} \left| \frac{1}{n} \sum_{i=1}^n W_i W_i^\tau K_h(T_i - t) - f(t)\Gamma(t) \right| \\ &\le \max_{1\le l\le l_n} \sup_{t\in\Delta_l} \left| \frac{1}{n} \sum_{i=1}^n W_i W_i^\tau K_h(T_i - t) - \frac{1}{n} \sum_{i=1}^n W_i W_i^\tau K_h(T_i - t_l) \right| \\ &\quad + \max_{1\le l\le l_n} \left| \frac{1}{n} \sum_{i=1}^n W_i W_i^\tau K_h(T_i - t_l) - f(t_l)\Gamma(t_l) \right| \\ &\quad + \max_{1\le l\le l_n} \sup_{t\in\Delta_l} \left| f(t_l)\Gamma(t_l) - f(t)\Gamma(t) \right| \\ &:= I_1 + I_2 + I_3. \end{aligned}$$

Therefore, to prove (6.4), it is sufficient to show that $I_k = O_p(c_n),\ \ k = 1, 2, 3$.

Using the Lipschitz continuity of $K(\cdot)$, we have $|K_h(T_i - t) - K_h(T_i - t_l)| \le \frac{C_1}{h^2}|t - t_l| I(|T_i - t_l| \le C_2 h) \le \frac{C_1 r_n}{h^2} I(|T_i - t_l| \le C_2 h)$. Therefore, the (k_1, k_2) component in I_1, $1 \le k_1 \le k_2 \le p$, can be written as

$$\begin{aligned} &\max_{1\le l\le l_n} \sup_{t\in\Delta_l} \left| \frac{1}{n} \sum_{i=1}^n W_{ik_1} W_{ik_2} [K_h(T_i - t) - K_h(T_i - t_l)] \right| \\ &\le \frac{C_1 r_n}{nh^2} \max_{1\le l\le l_n} \left| \sum_{i=1}^n |W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h) \right. \end{aligned}$$

$$- \sum_{i=1}^{n} E|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h)\Big|$$
$$+ \frac{C_1 r_n}{nh^2} \max_{1 \le l \le l_n} \sum_{i=1}^{n} E|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h) := I_{11} + I_{12}.$$

For I_{11}, applying Lemmas 6.1 and 6.2 we have

$$P\Big(\frac{C_1 r_n}{h^2} \max_{1 \le l \le l_n} \Big|\frac{1}{n} \sum_{i=1}^{n} [|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h)$$
$$- E|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h)]\Big| \ge C_0 \sqrt{\frac{\log n}{nh}}\Big)$$
$$\le \sum_{l=1}^{l_n} \Big\{4 \exp\Big[\frac{-\frac{1}{16} C_0^2 nh \log n n^{-2/\delta}}{\frac{n}{m} D_m + \frac{1}{3} C_0 \sqrt{nh \log n} n^{-1/\delta} C_1 m}\Big] + 32 \frac{C_1}{C_0 \sqrt{nh \log n} n^{-1/\delta}} n\alpha(m)\Big\},$$

where $D_m = \max_{1 \le j \le 2m} E(h \sum_{i=1}^{j} [|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h) - E|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h)])^2 n^{-2/\delta} \le \frac{C_2 mh}{n^{2/\delta}}$. Taking $m = [\frac{n^{1-1/\delta} h}{C_0 \sqrt{nh \log n}}]$, we have

$$P\Big(\max_{1 \le l \le l_n} \Big|\frac{1}{n} \sum_{i=1}^{n} [|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h)$$
$$- E|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h)]\Big| \ge C_0 \sqrt{\frac{\log n}{nh}}\Big)$$
$$\le l_n \Big\{\frac{4}{n} + C_1 \frac{C_1 n^{1+1/\delta}}{\sqrt{nh \log n}} \alpha(m)\Big\} \le \frac{C_0}{n} l_n \to 0. \tag{6.5}$$

On the other hand, we have $E|W_{ik_1} W_{ik_2}| I(|T_i - t_l| \le C_2 h) = O(h)$. Therefore $I_{12} = O(\sqrt{\frac{\log n}{nh}})$. Together with (6.5), one can derive $I_1 = O_p(C_n)$. One can rewrite I_2 as

$$I_2 \le \max_{1 \le l \le l_n} \Big|\frac{1}{n} \sum_{i=1}^{n} [W_i W_i^\tau - \Gamma(T_i)] K_h(T_i - t_l)\Big|$$
$$+ \max_{1 \le l \le l_n} \Big|\frac{1}{n} \sum_{i=1}^{n} \Gamma(T_i) K_h(T_i - t_l) - E\Gamma(T_i) K_h(T_i - t_l)\Big|$$
$$+ \max_{1 \le l \le l_n} |E\Gamma(T_i) K_h(T_i - t_l) - f(t_l)\Gamma(t_l)| := I_{21} + I_{22} + I_{23}.$$

By the same technique used in proving (6.5), we have $I_{21} = O_p\left(\sqrt{\frac{\log n}{nh}}\right)$, $I_{22} = O_p\left(\sqrt{\frac{\log n}{nh}}\right)$. Using Taylor's expansion, we have $I_{23} = O(h^2)$. From (A1), we have

$$I_3 = \max_{1\le l\le l_n} \sup_{t\in\Delta_l} |f(t_l)\Gamma(t_l) - f(t)\Gamma(t)| \le C_1 r_n^2 + C_2 r_n = O\Big(\sqrt{\frac{\log n}{nh}}\Big).$$

Thus, (6.4) is proved, which completes the proof of this lemma. □

Lemma 6.7 *Suppose* (A1)–(A3), (A5) *and* (A6) *are satisfied, then* $\frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i\tilde{\xi}_i^{\tau} \stackrel{P}{\to} \Sigma_e + EX_1X_1^\tau - E[\Phi^\tau(T_1)\Gamma^{-1}(T_1)\Phi(T_1)]$.

Proof From the definition $\tilde{\xi}_i^{\tau} = \xi_i^\tau - S_i\boldsymbol{\xi}$ and (1.1), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i\tilde{\xi}_i^{\tau} &= \frac{1}{n}\sum_{i=1}^n (X_i^\tau - S_i\mathbf{X})^\tau (X_i^\tau - S_i\mathbf{X}) + \frac{1}{n}\sum_{i=1}^n (e_i^\tau - S_i\mathbf{e})^\tau (X_i^\tau - S_i\mathbf{X})\\ &\quad + \frac{1}{n}\sum_{i=1}^n (X_i^\tau - S_i\mathbf{X})^\tau (e_i^\tau - S_i\mathbf{e}) + \frac{1}{n}\sum_{i=1}^n (e_i^\tau - S_i\mathbf{e})^\tau (e_i^\tau - S_i\mathbf{e}),\end{aligned}$$

where $S_i = (W_i^\tau,\ 0)(D_{T_i}^\tau \omega_{T_i} D_{T_i})^{-1} D_{T_i}^\tau \omega_{T_i}$. By (6.1) and (6.2) in Lemma 6.6, we have

$$\begin{aligned}S_i\mathbf{X} &= (W_i^\tau,\ 0)(D_{T_i}^\tau \omega_{T_i} D_{T_i})^{-1} D_{T_i}^\tau \omega_{T_i}\mathbf{X}\\ &= (W_i^\tau,\ 0)\Big\{[nf(T_i)\Gamma(T_i)]^{-1} \otimes \frac{1}{\mu_2}\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\Big\}\\ &\quad \times \Big\{n\Phi(T_i)f(T_i) \otimes \begin{pmatrix}1\\0\end{pmatrix}\{1+O_p(c_n)\}\Big\}\\ &= (W_i^\tau,\ 0)\Big\{[nf(T_i)\Gamma(T_i)]^{-1}[n\Phi(T_i)f(T_i)] \otimes \frac{1}{\mu_2}\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\{1+O_p(c_n)\}\Big\}\\ &= (W_i^\tau,\ 0)\Big\{\Gamma^{-1}(T_i)\Phi(T_i) \otimes \frac{1}{\mu_2}\begin{pmatrix}1\\0\end{pmatrix}\{1+O_p(c_n)\}\Big\}\\ &= W_i^\tau \Gamma^{-1}(T_i)\Phi(T_i)\{1+O_p(c_n)\}. \end{aligned} \tag{6.6}$$

Similarly, using the approaches above and those in the proof of (6.1) and (6.2), we have

$$S_i\mathbf{e} = W_i^\tau\Gamma^{-1}(T_i)E(W_ie_i^\tau|T_i)\{1+O_p(c_n)\} = 0. \tag{6.7}$$

From (6.6) and using Lemma 6.5, it follows that

$$\frac{1}{n}\sum_{i=1}^n (X_i^\tau - S_i\mathbf{X})^\tau (X_i^\tau - S_i\mathbf{X}) \stackrel{P}{\to} EX_1X_1^\tau - E[\Phi^\tau(T_1)\Gamma^{-1}(T_1)\Phi(T_1)].$$

Similarly $\frac{1}{n}\sum_{i=1}^n (e_i^\tau - S_i\mathbf{e})^\tau (X_i^\tau - S_i\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n e_i(X_i^\tau - W_i^\tau\Gamma^{-1}(T_i)\Phi(T_i))\{1 + O_p(c_n)\} \stackrel{P}{\to} 0$. According to (6.7), we have $\frac{1}{n}\sum_{i=1}^n (e_i^\tau - S_i\mathbf{e})^\tau (e_i^\tau - S_i\mathbf{e}) = \frac{1}{n}\sum_{i=1}^n e_ie_i^\tau \stackrel{a.s.}{\to} \Sigma_e$. Thus the conclusion is proved. □

Lemma 6.8 *Suppose* (A1)–(A6) *are satisfied, then* $\sum_{i=1}^{n}\tilde{\xi}_i\tilde{M}_i = o_p(\sqrt{n})$, *where* $\tilde{M}_i = M_i - S_i M$ *and* $M_i = W_i^\tau a(T_i)$.

Proof According to the definition, we have

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i\tilde{M}_i = \frac{1}{n}\sum_{i=1}^{n}(X_i^\tau - S_i\mathbf{X})^\tau(M_i^\tau - S_i M) + \frac{1}{n}\sum_{i=1}^{n}(e_i^\tau - S_i\mathbf{e})^\tau(M_i^\tau - S_i M). \tag{6.8}$$

Note that $D_t^\tau\omega_t M = \begin{pmatrix}\sum_{i=1}^{n} W_i W_i^\tau a(T_i)K_h(T_i-t)\\ \sum_{i=1}^{n} W_i W_i^\tau a(T_i)\frac{T_i-t}{h}K_h(T_i-t)\end{pmatrix}$. Using the similar techniques in the proof of Lemma 6.6, one can easily check that $D_t^\tau\omega_t M = n\Gamma(t)f(t)a(t)\otimes\begin{pmatrix}1\\0\end{pmatrix}\{1+O_p(c_n)\}$. Therefore $S_i M = W_i^\tau a(T_i)\{1+O_p(c_n)\}$, furthermore,

$$\tilde{M}_i = M_i - S_i M = W_i^\tau a(T_i)O_p(c_n). \tag{6.9}$$

Then, from (6.6) and law of large numbers for stationary α-mixing sequences, one can obtain

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(X_i^\tau - S_i\mathbf{X})^\tau(M_i^\tau - S_i M)\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}[X_i^\tau - W_i^\tau\Gamma^{-1}(T_i)\Phi(T_i) - W_i^\tau\Gamma^{-1}(T_i)\Phi(T_i)O_p(c_n)]^\tau W_i^\tau a(T_i)O_p(c_n)\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}X_i W_i^t a(T_i)O_p(c_n) - \frac{1}{n}\sum_{i=1}^{n}\Phi^\tau(T_i)\Gamma^{-1}(T_i)W_i W_i^\tau a(T_i)O_p(c_n)\\
&\qquad - \frac{1}{n}\sum_{i=1}^{n}\Phi^\tau(T_i)\Gamma^{-1}(T_i)W_i W_i^\tau a(T_i)O_p(c_n^2)\\
&\quad= E[\Phi^\tau(T_1)a(T_1)]O_p(c_n^2).
\end{aligned} \tag{6.10}$$

Similarly with (6.7), we have $\frac{1}{n}\sum_{i=1}^{n}(e_i^\tau - S_i\mathbf{e})^\tau(M_i^\tau - S_i M)\xrightarrow{P}0$, which, together with (6.8) and (6.10), yields that $\sum_{i=1}^{n}\tilde{\xi}_i\tilde{M}_i = O_p(nc_n^2) = o_p(\sqrt{n})$. □

Lemma 6.9 (i) *Suppose* (A1)–(A6) *are satisfied, then*

$$\sqrt{n}(\hat{\beta}_n - \beta)\xrightarrow{\mathcal{D}}N(0,\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}),$$

where $\Sigma_1 = E(X_1X_1^\tau) - E[\Phi^\tau(T_1)\Gamma^{-1}(T_1)\Phi(T_1)]$, $\Phi(T_1) = E(W_1X_1^\tau|T_1)$, $\Gamma(T_1) = E(W_1W_1^\tau|T_1)$ $\Sigma_2 = \lim_{n\to\infty}Var\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\xi_i - \Psi^\tau(T_i)\Gamma^{-1}(T_i)W_i]$ $[\epsilon_i - e_i^\tau\beta]\}$. *Further,* $\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}$ *is a consistent estimator of* $\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$, *where*

$\hat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i\tilde{\xi}_i^\tau - \Sigma_e$, $\hat{\Sigma}_2 = \frac{1}{n}\Big\{\sum_{i=1}^n[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\hat{\beta}_n)] + \Sigma_e\hat{\beta}_n\Big\}^{\otimes 2}$, *here* $C^{\otimes 2}$ *means* CC^τ.

(ii) *Suppose* (A1)–(A6) *are satisfied, then* $\sqrt{n}(\hat{\beta}_J - \beta) = \sqrt{n}(\hat{\beta}_n - \beta) + o_p(1)$.

Proof (i) Let $\sum_{i=1}^n \tilde{\xi}_i\tilde{\xi}_i^\tau - n\Sigma_e = A$, then $\hat{\beta}_n = A^{-1}\sum_{i=1}^n \tilde{\xi}_i\tilde{Y}_i^\tau$. Write

$$\hat{\beta}_n - \beta = A^{-1}n\Sigma_e\beta + A^{-1}\sum_{i=1}^n \tilde{\xi}_i(\tilde{Y}_i^\tau - \tilde{\xi}_i^\tau\beta). \tag{6.11}$$

From Lemma 6.7, we have $A^{-1} = O(\frac{1}{n})$. According to the definition and (1.1), we write

$$\frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i(\tilde{Y}_i^\tau - \tilde{\xi}_i^\tau\beta) = \frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i(M_i - S_iM) + \frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i(\epsilon_i - S_i\epsilon)$$
$$- \frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i(e_i^\tau - S_i\mathbf{e}^\tau)\beta. \tag{6.12}$$

From (6.6) and (6.7), we have

$$\frac{1}{n}\sum_{i=1}^n \tilde{\xi}_i(e_i^\tau - S_i\mathbf{e}^\tau)\beta = \frac{1}{n}\sum_{i=1}^n[\xi_i - \Phi^\tau(T_i)\Gamma^{-1}(T_i)W_i]e_i^\tau\beta + o_p\Big(\frac{1}{\sqrt{n}}\Big). \tag{6.13}$$

Similar to the proof of (6.2) in Lemma 6.6, one can easily check that $D_t^\tau\omega_t\epsilon = n\mathbf{1}_{2q}\otimes\binom{1}{0}O_p\Big(\sqrt{\frac{\log n}{nh}}\Big)$. Together with (6.1), (A1) and (A2), we have

$$\begin{aligned} S_i\epsilon &= (W_i^\tau,\ \ 0)(D_{T_i}^\tau\omega_{T_i}D_{T_i})^{-1}D_{T_i}\omega_{T_i}\epsilon \\ &= (W_i^\tau,\ \ 0)\Big\{[nf(T_i)\Gamma(T_i)]^{-1}\otimes\frac{1}{\mu_2}\begin{pmatrix}\mu_2 & 0\\ 0 & 1\end{pmatrix}\Big\}\Big\{n\mathbf{1}_{2q}\otimes\binom{1}{0}\Big\}O_p\Big(\sqrt{\frac{\log n}{nh}}\Big) \\ &= W_i^\tau\mathbf{1}_q[f(T_i)\Gamma(T_i)]^{-1}O_p\Big(\sqrt{\frac{\log n}{nh}}\Big) = W_i^\tau\mathbf{1}_qO_p\Big(\sqrt{\frac{\log n}{nh}}\Big). \end{aligned} \tag{6.14}$$

Therefore

$$\begin{aligned} \sum_{i=1}^n \tilde{\xi}_i(\epsilon_i - S_i\epsilon) &= \sum_{i=1}^n e_i\epsilon_i + \sum_{i=1}^n(X_i - \Phi^\tau(T_i)\Gamma^{-1}(T_i)W_i)\epsilon_i + o_p(\sqrt{n}) \\ &= \sum_{i=1}^n(\xi_i - \Phi^\tau(T_i)\Gamma^{-1}(T_i)W_i)\epsilon_i + o_p(\sqrt{n}). \end{aligned} \tag{6.15}$$

Combining (6.11)–(6.15) and Lemma 6.8, we have

$$\sqrt{n}(\hat{\beta}_n-\beta)=\left(\frac{A}{n}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\Sigma_e\beta+[\xi_i-\Phi^\tau(T_i)\Gamma^{-1}(T_i)W_i][\epsilon_i-e^\tau\beta]\right\}+o_p(1).$$

Let $\eta_i = \Sigma_e\beta + [\xi_i - \Phi^\tau(T_i)\Gamma^{-1}(T_i)W_i][\epsilon_i - e^\tau\beta]$. Obviously, $\{\eta_i, i \geq 1\}$ is an α-mixing sequence with $E\eta_i = 0$ and $E|\eta_i|^\delta < \infty$ for $\delta > 4$. Applying Lemma 6.3, one can complete the proof of (i).

(ii) To prove $\sqrt{n}(\hat{\beta}_J - \beta) = \sqrt{n}(\hat{\beta}_n - \beta) + o_p(1)$, it is sufficient to prove $\hat{\beta}_J = \hat{\beta}_n + o_p(\frac{1}{\sqrt{n}})$.

Note that $\hat{\beta}_J = \hat{\beta}_n + \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\beta}_n - \hat{\beta}_{n,-i})$. Therefore, we only need to prove that

$$\sqrt{n}\sum_{i=1}^{n}(\hat{\beta}_n - \hat{\beta}_{n,-i}) = o_p(1). \tag{6.16}$$

From the definition,

$$\hat{\beta}_n - \hat{\beta}_{n,-i}=\left[\sum_{i=1}^{n}\tilde{\xi}_i\tilde{\xi}_i^\tau - n\Sigma_e\right]^{-1}\sum_{i=1}^{n}\tilde{\xi}_i\tilde{Y}_i-\left[\sum_{j\neq i}^{n}\tilde{\xi}_j\tilde{\xi}_j^\tau-(n-1)\Sigma_e\right]^{-1}\sum_{j\neq i}^{n}\tilde{\xi}_j\tilde{Y}_j.$$

Using the fact [see Theorem 11.2.3 in Golub and Van Loan (1996)] $(A+B)^{-1} = A^{-1} - A^{-1}BA^{-1} - A^{-1}B\sum_{k=1}^{\infty}C^kA^{-1}$, where A is a nonsingular matrix, and $C = -A^{-1}B$. We write

$$\begin{aligned}&\left[\sum_{j\neq i}\tilde{\xi}_j\tilde{\xi}_j^\tau - (n-1)\Sigma_e\right]^{-1}\\ &\quad=\left[\sum_{j\neq i}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e\right]^{-1}-\left[\sum_{j\neq i}\tilde{\xi}_j\tilde{\xi}_j^\tau-n\Sigma_e\right]^{-1}\Sigma_e\left[\sum_{j\neq i}\tilde{\xi}_j\tilde{\xi}_j^\tau-n\Sigma_e\right]^{-1}-D,\end{aligned} \tag{6.17}$$

where $D = A^{-1}B\sum_{k=1}^{\infty}C^kA^{-1}$, $A = [\sum_{j\neq i}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e]$, $B = \Sigma_e$, $C = -A^{-1}B$.

Applying Lemma 6.4, we write

$$\begin{aligned}\left[\sum_{j\neq i}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e\right]^{-1} &= \left[\sum_{j=1}^{n}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e - \tilde{\xi}_i\tilde{\xi}_i^\tau\right]^{-1}\\ &=\left[\sum_{j=1}^{n}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e\right]^{-1}+\frac{[\sum_{j=1}^{n}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e]^{-1}\tilde{\xi}_i\tilde{\xi}_i^\tau[\sum_{j=1}^{n}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e]^{-1}}{1-\tilde{\xi}_i^\tau[\sum_{j=1}^{n}\tilde{\xi}_j\tilde{\xi}_j^\tau - n\Sigma_e]^{-1}\tilde{\xi}_i}.\end{aligned} \tag{6.18}$$

Let $A = [\sum_{j=1}^{n} \tilde{\xi}_j \tilde{\xi}_j^{\tau} - n\Sigma_e]$, the same as in the proof of Lemma 6.9 (i). Then combining (6.17), (6.18) and the definitions of $\hat{\beta}_n$ and $\hat{\beta}_{n,-i}$ and noting that $\sum_{i=1}^{n}[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^{\tau}\hat{\beta}_n) + \Sigma_e\hat{\beta}_n] = 0$, we write

$$\begin{aligned}
&\sum_{i=1}^{n}(\hat{\beta}_n - \hat{\beta}_{n,-i}) \\
&= A^{-1}\sum_{i=1}^{n}\frac{v_i\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^{\tau}\hat{\beta}_n) + \Sigma_e\hat{\beta}_n}{1-v_i} - A^{-1}\sum_{i=1}^{n}\frac{r_i[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^{\tau}\hat{\beta}_n) + \Sigma_e\hat{\beta}_n]}{(1-v_i)^2} \\
&\quad - A^{-1}\Sigma_e A^{-1}\sum_{i=1}^{n}\frac{\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^{\tau}\hat{\beta}_n) + \Sigma_e\hat{\beta}_n}{1-v_i} - A^{-1}\sum_{i=1}^{n}\frac{v_i}{1-v_i}\Sigma_e\hat{\beta}_n \\
&\quad + A^{-1}\Sigma_e A^{-1}\sum_{i=1}^{n}\frac{1}{1-v_i}\Sigma_e\hat{\beta}_n \\
&\quad + A^{-1}\sum_{i=1}^{n}r_i\frac{\Sigma_e\hat{\beta}_n}{(1-v_i)^2} + A^{-1}\sum_{i=1}^{n}\frac{\tilde{\xi}_i\tilde{\xi}_i^{\tau}A^{-1}\Sigma_e\hat{\beta}_n}{1-v_i} + D\sum_{i=1}^{n}\sum_{j\neq i}\tilde{\xi}_j\tilde{Y}_j \\
&:= A^{-1}\sum_{i=1}^{7}I_i + D\sum_{i=1}^{n}\sum_{j\neq i}\tilde{\xi}_j\tilde{Y}_j, \qquad (6.19)
\end{aligned}$$

where $v_i = \tilde{\xi}_i^{\tau}A^{-1}\tilde{\xi}_i$, $r_i = \tilde{\xi}_i^{\tau}A^{-1}\Sigma_e A^{-1}\tilde{\xi}_i$. By Lemma 6.7 and (A3), we have $v_i = O_p(n^{-1})$ and $r_i = O_p(n^{-2})$. Therefore, to prove (6.16), it is sufficient to prove that

$$I_i = o_p(\sqrt{n}),\quad i = 1, 2, \cdots, 7 \ \text{ and } \ D\sum_{i=1}^{n}\sum_{j\neq i}\tilde{\xi}_j\tilde{Y}_j = o_p(\frac{1}{\sqrt{n}}).$$

First, we deal with I_1. Since

$$\begin{aligned}
\left|\sum_{i=1}^{n}\frac{v_i}{1-v_i}[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^{\tau}\hat{\beta}_n) + \Sigma_e\hat{\beta}_n]\right| &\leq \sqrt{n}(\max_{1\leq i\leq n} v_i^2)^{1/2} \\
&\left(\sum_{i=1}^{n}[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^{\tau}\hat{\beta}_n) + \Sigma_e\hat{\beta}_n]^2\right)^{1/2},
\end{aligned}$$

to prove the desired result, one needs only to show that

$$\left(\max_{1\leq i\leq n} v_i^2\sum_{i=1}^{n}[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^{\tau}\hat{\beta}_n) + \Sigma_e\hat{\beta}_n]^2\right)^{1/2} = o_p(1).$$

In fact, from $\max_{1\le i\le n}|v_i| = o(n^{-3/4})$ $a.s.$ by the proof of Lemma 3 in Owen (1990), and Lemma 6.11, it follows that $\frac{1}{\sqrt{n}}I_1 = o_p(1)$. Similarly $\frac{1}{\sqrt{n}}I_2 = o_p(1)$, $\frac{1}{\sqrt{n}}I_3 = o_p(1)$.

Meanwhile, $\|\frac{1}{\sqrt{n}}I_{n4}\| = \frac{n}{\sqrt{n}}O_p(\frac{1}{n}) \to 0$. Similarly, we have

$$\frac{1}{\sqrt{n}}I_5 = o_p(1),\ \frac{1}{\sqrt{n}}I_6 = o_p(1),\ \frac{1}{\sqrt{n}}I_7 = o_p(1).$$

Recall the definition of A, B, C, D and Lemma 6.7, we have $A^{-1} = O(\frac{1}{n}), C = O(\frac{1}{n})$ and

$$D = A^{-1}B(CA^{-1} + C^2A^{-1} + C^3A^{-1} + \cdots) = \frac{1}{n^3} + \frac{1}{n^4} + \cdots = O\Big(\frac{1}{n^3}\Big).$$

Therefore, by (A3), one can easily obtain that $\sqrt{n}D\sum_{i=1}^n\sum_{j\ne i}^n \tilde{\xi}_j\tilde{Y}_j = \sqrt{n}O\Big(\frac{1}{n^3}\Big)n^2O_p(1) \to 0$. □

Lemma 6.10 *Suppose* (A3) *and* (A6) *are satisfied, then* $\frac{1}{n}\sum_{i=1}^n \epsilon_i W_{ik} = o(n^{-1/4})$ $a.s.$ *for* $1\le k\le p$.

Proof Following the proof of Lemma 2 in Hong and Cheng (1994) under the independent case, using Lemmas 6.1 and 6.2, it is not difficult to prove this lemma. □

Lemma 6.11 *Suppose* (A1)–(A3), (A5) *and* (A6) *are satisfied, then* $\frac{1}{n}\sum_{i=1}^n[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\hat{\beta}_n) + \Sigma_e\hat{\beta}_n][\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\hat{\beta}_n) + \Sigma_e\hat{\beta}_n]^\tau \xrightarrow{P} \Sigma_3$ *and* $\max_{1\le i\le n}\|\hat{\beta}_n - \hat{\beta}_{n,-i}\| = O_p(n^{-1})$, *where* $\Sigma_3 = (\Sigma_1 + \Sigma_e)(\sigma^2 + \beta^\tau\Sigma_e\beta) - \Sigma_e\beta\beta^\tau\Sigma_e$.

Proof (i) Write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\hat{\beta}_n) + \Sigma_e\hat{\beta}_n][\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\hat{\beta}_n) + \Sigma_e\hat{\beta}_n]^\tau \\
&\quad = \frac{1}{n}\sum_{i=1}^n[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}^\tau\beta)][\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}^\tau\beta)]^\tau + \frac{1}{n}\sum_{i=1}^n\tilde{\xi}_i\tilde{\xi}_i^\tau(\hat{\beta}_n - \beta)(\hat{\beta}_n - \beta)^\tau\tilde{\xi}_i\tilde{\xi}_i^\tau \\
&\qquad + \frac{1}{n}\sum_{i=1}^n\Sigma_e\hat{\beta}_n\hat{\beta}_n^\tau\Sigma_e - \frac{1}{n}\sum_{i=1}^n\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\beta)\tilde{\xi}_i^\tau\tilde{\xi}_i^\tau(\hat{\beta}_n - \beta) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\beta)\hat{\beta}_n^\tau\Sigma_e - frac1n\sum_{i=1}^n\tilde{\xi}_i\tilde{\xi}_i^\tau(\hat{\beta}_n - \beta)\hat{\beta}_n^\tau\Sigma_e \\
&\qquad - \frac{1}{n}\sum_{i=1}^n[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\beta)\tilde{\xi}_i^\tau\tilde{\xi}_i^\tau(\hat{\beta}_n - \beta)]^\tau + \frac{1}{n}\sum_{i=1}^n[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau\beta)\hat{\beta}_n^\tau\Sigma_e]^\tau \\
&\qquad - \frac{1}{n}\sum_{i=1}^n[\tilde{\xi}_i\tilde{\xi}_i^\tau(\hat{\beta}_n - \beta)\hat{\beta}_n^\tau\Sigma_e]^\tau.
\end{aligned} \tag{6.20}$$

First, we evaluate the cross terms. By Lemmas 6.9 and 6.5, (A2) and (A3), we have

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i\tilde{\xi}_i^\tau(\hat{\beta}_n-\beta)\hat{\beta}_n^\tau\Sigma_e\right\| = \left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i\tilde{\xi}_i^\tau\right\|\left\|\hat{\beta}_n-\beta\right\|\left\|\hat{\beta}_n^\tau\Sigma_e\right\| = O_p(n^{-1/2}) \to 0.$$

Similarly $\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}_i^\tau\beta)\tilde{\xi}_i^\tau\tilde{\xi}_i^\tau(\hat{\beta}_n-\beta)\| \xrightarrow{P} 0$. Note that $\sum_{i=1}^{n}[\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}_i^\tau\hat{\beta}_n)+\Sigma_e\hat{\beta}_n]=0$, with Lemma 6.7 we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}_i^\tau\beta)\hat{\beta}_n^\tau\Sigma_e &= -\Sigma_e\hat{\beta}_n\hat{\beta}_n^\tau\Sigma_e+\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i\tilde{\xi}_i^\tau(\hat{\beta}_n-\beta)\hat{\beta}_n^\tau\Sigma_e\\ &= -\Sigma_e\hat{\beta}_n\hat{\beta}_n^\tau\Sigma_e+(\Sigma_e+\Sigma_1)(\hat{\beta}_n-\beta)\hat{\beta}_n^\tau\Sigma_e \xrightarrow{P} -\Sigma_e\beta\beta^\tau\Sigma_e.\end{aligned}$$

Therefore, one can write (6.20) as

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}[\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}_i^\tau\hat{\beta}_n)+\Sigma_e\hat{\beta}_n]\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}_i^\tau\hat{\beta}_n)+\Sigma_e\hat{\beta}_n]^\tau\\ &=\frac{1}{n}\sum_{i=1}^{n}[\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}^\tau\beta)][\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}^\tau\beta)]^\tau+\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i\tilde{\xi}_i^\tau(\hat{\beta}_n-\beta)(\hat{\beta}_n-\beta)^\tau\tilde{\xi}_i\tilde{\xi}_i^\tau\\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\Sigma_e\hat{\beta}_n\hat{\beta}_n^\tau\Sigma_e-2\Sigma_e\beta\beta^\tau\Sigma_e\\ &:= H_1+H_2+H_3-2\Sigma_e\beta\beta^\tau\Sigma_e.\end{aligned}$$

On applying Lemma 6.5 and (6.6) we have

$$\begin{aligned}H_1 &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i\tilde{\xi}_i^\tau(\epsilon_i-e_i^\tau\beta)^2=\frac{1}{n}\sum_{i=1}^{n}(X_i^\tau-W_i^\tau\Gamma^{-1}(T_i)\Phi(T_i)\\ &\qquad (1+O_p(c_n))+e_i^\tau)^2(\epsilon_i-e_i^\tau\beta)^2\\ &\xrightarrow{P} E[X_1^\tau-W_1^\tau\Gamma^{-1}(T_1)\Phi(T_1)+e_1^\tau]^2(\epsilon_1-e_1^\tau\beta)^2=(\sigma^2+\beta^\tau\Sigma_e\beta)(\Sigma_1+\Sigma_e).\end{aligned}$$

With $\max_{1\le i\le n}\|\tilde{\xi}_i\|=o(n^{1/2\delta})$, $\|\hat{\beta}_n-\hat{\beta}\|=O_p(n^{-1/2})$, and Lemma 6.7, one can derive that $H_2\to 0$, $H_3\to\Sigma_e\beta\beta^\tau\Sigma_e$. Hence, the first conclusion is verified.

Similar to the derivation of (6.19), one can write

$$\begin{aligned}\hat{\beta}_n-\hat{\beta}_{n,-i} &= A^{-1}\frac{\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}_i^\tau\hat{\beta}_n)+\Sigma_e\hat{\beta}_n}{1-v_i}-A^{-1}\frac{v_i}{1-v_i}\Sigma_e\hat{\beta}_n\\ &\quad -A^{-1}\Sigma_eA^{-1}\frac{\tilde{\xi}_i(\tilde{Y}_i-\tilde{\xi}_i^\tau\hat{\beta}_n)+\Sigma_e\hat{\beta}_n}{1-v_i}\\ &\quad +A^{-1}\Sigma_eA^{-1}\frac{\Sigma_e\hat{\beta}_n}{1-v_i}+A^{-1}\frac{\tilde{\xi}_i\tilde{\xi}_i^\tau A^{-1}\Sigma_e\hat{\beta}_n}{1-v_i}\end{aligned}$$

$$- A^{-1} r_i \frac{\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \Sigma_e \hat{\beta}_n}{(1 - v_i)^2} \\ + A^{-1} r_i \frac{\Sigma_e \hat{\beta}_n}{(1 - v_i)^2} + D \sum_{j \neq i} \tilde{\xi}_j \tilde{Y}_j := \sum_{k=1}^{8} a_{ki},$$

where $v_i = \tilde{\xi}_i^\tau A^{-1} \tilde{\xi}_i$, $r_i = \tilde{\xi}_i^\tau A^{-1} \Sigma_e A^{-1} \tilde{\xi}_i$. Then, it is sufficient to show that

$$\max_{1 \leq i \leq n} \|a_{ki}\| = O_p(n^{-1}), \quad k = 1, 2, \cdots, 8.$$

For a_{1i}, since $E[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \Sigma_e \hat{\beta}_n] = 0$, $E\|\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \Sigma_e \hat{\beta}_n\|^\delta < \infty$ and $\max_{1 \leq i \leq n} |v_i| = o(n^{-3/4})$ $a.s.$, we have $\max_{1 \leq i \leq n} \|\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \Sigma_e \hat{\beta}_n\| = O_p(1)$. Therefore, $\max_{1 \leq i \leq n} \|a_{1i}\| = O_p(n^{-1})$. It is easy to see that

$$\frac{1}{n} \sum_{i=1}^{n} a_{i3}^2 = \frac{1}{n} \sum_{i=1}^{n} [\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \Sigma_e \hat{\beta}_n][\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^\tau \hat{\beta}_n) + \Sigma_e \hat{\beta}_n]^\tau O(n^{-4}) = O(n^{-4}),$$

which implies $\frac{n^2 \max_{1 \leq i \leq n} \|a_{i3}\|}{\sqrt{n}} \to 0$. Then $\max_{1 \leq i \leq n} \|a_{i3}\| = o_p(n^{-3/2})$. Similarly, $\max_{1 \leq i \leq n} \|a_{6i}\| = o_p(n^{-3/2})$.

From $\max_{1 \leq i \leq n} |v_i| = o(1)$ $a.s.$, $\max_{1 \leq i \leq n} |r_i| = o(n^{-1})$ $a.s.$ and $\max_{1 \leq i \leq n} \|\tilde{\xi}_i\| = o(n^{1/2\delta})$ $a.s.$, it is easy to show that $\max_{1 \leq i \leq n} \|a_{2i}\| = o(n^{-1})$, $\max_{1 \leq i \leq n} \|a_{4i}\| = O(n^{-2})$, $\max_{1 \leq i \leq n} \|a_{5i}\| = o(n^{-1})$, $\max_{1 \leq i \leq n} \|a_{7i}\| = o(n^{-2})$, $\max_{1 \leq i \leq n} \|a_{8i}\| = o(n^{-1})$.

Then the proof of the second conclusion is completed. □

References

Ahmad I, Leehalanon S, Li Q (2005) Efficient estimation of a semiparametric partially linear varying coefficient model. Ann Stat 33:258–283

Bravo F (2014) Varying coefficients partially linear models with randomly censored data. Ann Inst Stat Math 66:383–412

Doukhan P (1994) Mixing: properties and examples. Springer, New York

Fan GL, Xu HX, Liang HY (2012) Empirical likelihood inference for partially time-varying coefficient errors-in-variables models. Electron J Stat 6:1040–1058

Fan GL, Liang HY, Wang JF (2013) Statistical inference for partially linear time-varying coefficient errors-in-variables models. J Stat Plann Inference 143:505–519

Fan GL, Liang HY, Wang JF (2013) Empirical likelihood for heteroscedastic partially linear errors-in-variables model with α-mixing errors. Stat Pap 54:85–112

Fan J, Huang T (2005) Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Beroulli 11:1031–1057

Feng H, Peng L (2012) Jackknife empirical likelihood tests for distribution functions. J Stat Plan Inference 142:1571–1585

Feng S, Xue L (2014) Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition. Ann Inst Stat Math 66:121–140

Golub GH, Van Loan CF (1996) Matrix computations, 3rd edn. John Hopkins University Press, Baltimore

Gong Y, Peng L, Qi Y (2010) Smoothed jackknife empirical likelihood method for roc curve. J Multivar Anal 101:1520–1531

Hall P (1992) The bootstrap and edgeworth expansion. Springer, New York
Hall P, La Scala B (1990) Methodology and algorithms of empirical likelihood. Int Stat Rev 58:109–127
Hong S, Cheng P (1994) The convergence rate of estimation for parameter in a semiparametric model. Chin J Appl Probab Stat 10:62–71
Huang Z, Zhang R (2009) Empirical likelihood for nonparametric parts in semiparametric varying-coefficient partially linear models. Stat Probab Lett 79:1798–1808
Jing BY, Yuan J, Zhou W (2009) Jackknife empirical likelihood. J Am Stat Assoc 104:1224–1232
Liang H, Härdle W, Carroll RJ (1999) Estimation in a semiparametric partially linear errors-in-variables model. Ann Stat 27:1519–1535
Liang HY, Jing BY (2009) Asymptotic normality in partially linear models based on dependent errors. J Stat Plan Inference 139:1357–1371
Liang HY, Mammitzsch V, Steinebach J (2006) On a semiparametric regression model whose errors form a linear process with negatively associated innovations. Statistics 40:207–226
Liebscher E (2001) Estimation of the density and the regression function under mixing conditions. Stat Decis 19:9–26
Lin Z, Lu C (1996) Limit theory for mixing dependent random variables. Science Press, New York
Miao Y, Zhao F, Wang K, Chen Y (2013) Asymptotic normality and strong consistency of LS estimators in the EV regression model with NA errors. Stat Pap 54:193–206
Miller RG (1974) An unbalanced jackknife. Ann Stat 2:880–891
Owen AB (1988) Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75:237–249
Owen AB (1990) Empirical likelihood ratio confidence regions. Ann Stat 8:90–120
Peng L (2012) Approximate jackknife empirical likelihood method for estimating equations. Can J Stat 40:110–123
Peng L, Qi Y, Van Keilegom I (2012) Jackknife empirical likelihood method for copulas. Test 21:74–92
Shao QM (1993) Complete convergence for α-mixing sequences. Stat Probab Lett 16:279–287
Singh S, Jain K, Sharma S (2014) Replicated measurement error model under exact linear restrictions. Stat Pap 55:253–274
Wang X, Li G, Lin L (2011) Empirical likelihood inference for semiparametric varying-coefficient partially linear EV models. Metrika 73:171–185
Wei C, Luo Y, Wu X (2012) Empirical likelihood for partially linear additive errors-in-variables models. Stat Pap 53:485–496
Yang SC (2007) Maximal moment inequality for partial sums of strong mixing sequences and application. Acta Math Sin Engl Ser 23:1013–1024
You J, Chen G (2006) Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. J Multivar Anal 97:324–341
You J, Chen G (2007) Semiparametric generalized least squares estimation in partially linear regression models with correlated errors. J Stat Plan Inference 137:117–132
You J, Zhou X, Chen G (2005) Jackknifing in partially linear regression models with serially correlated errrors. J Multivar Anal 92:386–404
You J, Zhou Y (2006) Empirical likelihood for semiparametric varying-coefficient partially linear regression models. Stati Probab Lett 76:412–422
Zhang JJ, Liang HY (2012) Asymptotic normality of estimators in heteroscedastic semiparametric model with strong mixing errors. Commun Stat 41:2172–2201
Zhou H, You J, Zhou B (2010) Statistical inference for fixed-effects partially linear regression models with errors in variables. Stat Pap 51:629–650
Zi X, Zou C, Liu Y (2012) Two-sample empirical likelihood method for difference between coefficients in linear regression model. Stat Pap 53:83–93