Transformation models with cointegrated and deterministically trending regressors

Yingqian Lin[a] and Yundong Tu[b*]

[a] School of Economics, Shanghai University of Finance and Economics

[b] Guanghua School of Management and

Center for Statistical Science, Peking University

October 14, 2022

Abstract

This paper develops an asymptotic theory for a general transformation model with a time trend, stationary regressors, and unit root nonstationary regressors. This model extends that of Han (1987) to incorporate time trend and nonstationary regressors. When the transformation is specified as an identity function, the model reduces to the conventional cointegrating regression, possibly with a time trend and other stationary regressors, which has been studied in Phillips and Durlauf (1986) and Park and Phillips (1988, 1989). The limiting distributions of the extremum estimator of the transformation parameter and the plug-in estimators of other model parameters are found to critically depend upon the transformation function and the order of the time trend. Simulations demonstrate that the estimators perform well in finite samples.

JEL classification: C13, C22, C51.

Keywords: Cointegration; extreme estimation; nonlinear model; time trend; transformation model; unit root.

*Corresponding author. Address: Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. E-mail: yundong.tu@gsm.pku.edu.cn.

1 Introduction

The concept of cointegration has been proved important since the seminal work of Granger (1981) and Engle and Granger (1987), with numerous developments witnessed in both theoretical and empirical analysis during the past three decades. The generalization of this concept to nonlinear cointegration dates back to Granger (1991), and gives rise to a wide range of studies in economics and related fields. As put forward by Park (2006), the nonlinear cointegration models provide more flexibilities and allow for much more diverse types of relationships among integrated processes. Well-known applications of nonlinear cointegrations include the modeling of money demand functions, as demonstrated by Bae et al. (2006), Bae and De Jong (2007) and Kim and Kim (2012), and the environmental Kuznets Curve as illustrated by Chan and Wang (2015), Wang et al. (2018), and Lin et al. (2020), among others. The readers are referred to Granger and Teräsvirta (1993) and Teräsvirta et al. (2010) for more empirical illustrations.

The statistical foundation of the nonlinear cointegration theory has been built up by Professor Joon Park and his co-authors in a sequence of influential papers. In particular, Park and Phillips (1999, 2001) developed a general asymptotic theory of nonlinear regressions with integrated time series, which has been generalized later by Chan and Wang (2015). Park and Phillips (2000) studied the binary choice models with integrated regressors. Chang et al. (2001) extended the model of Park and Phillips (2001) to the multivariate case by accommodating a time trend and stationary regressors, as well as multiple I(1) regressors. Chang and Park (2003) considered the nonlinear index models driven by integrated processes. Chang et al. (2012) considered nonstationary logistic regression. In addition, Chang and Park (2011) and Chan and Wang (2015) considered the endogeneity in nonstationary nonlinear regressions. More recently, Lin and Tu (2021b) investigated transformed linear cointegration models with multiple unit root processes.

Besides the parametric nonlinear cointegration, the nonstationary semi/nonparametric models become increasingly popular in econometrics, due to their flexibility to characterize nonlinear cointegrated relationships. For instance, Phillips and Park (1998), Park and Hahn (1999) and Juhl (2005) developed early nonparametric asymptotic analyses with nonstationary data; Karlsen et al. (2007), Wang and Phillips (2009a,b, 2016), Wang (2015) and Linton and Wang (2016) studied nonparametric cointegration models with kernel estimation method; Cai et al. (2009), Xiao (2009), Gao and Phillips (2013), Hirukawa and Sakudo (2018) and Tu and Wang (2019, 2020) considered functional coefficient cointegration models; Dong et al. (2016) considered the (partially) single index integrated model; Phillips et al. (2017) considered smooth structural changes in cointegration models; Dong and Linton (2018) proposed an additive nonparametric regression with time variable, nonstationary and stationary variables, while Dong et al. (2021) proposed a weighted sieve estimator for the more general nonparametric setting. Lin et al. (2020) proposed a double-nonlinear cointegration model in which a monotonic transformation of the dependent variable and a nonparametric transformation of a unit root regressor are cointegrated. In addition, a variety of specification tests have been developed in nonlinear cointegration models. See, for example, Wang and Phillips (2012), Wang et al. (2018) and Dong and Gao (2018) for specification tests in cointegrations with a univariate nonstationary regressor, Kasparis and Phillips (2012) for dynamic misspecification tests, and Kasparis et al. (2015) for inferences in nonparametric predictive regressions, Dong et al. (2017) and Tu et al. (2021) for specification tests in cointegrations with stationary covariates, among others. Moreover, Phillips (2009) and Tu and Wang (2021) studied spurious regressions in nonparametric regression and functional coefficient regressions with integrated processes.

This paper contributes to the above growing literature by investigating a new transformation nonlinear cointegration model, where a monotonic nonlinear transformation of the dependent variable is cointegrated with the multivariate unit root regressors, the time trend and the stationary regressors. Such a transformation model extends the model of Han (1987) to the case with a time trend and nonstationary regressors. In the special case that the transformation function becomes identity, the proposed model degenerates to the conventional linear cointegration model (possibly with stationary regressors and time trend) studied by Phillips and Durlauf (1986) and Park and Phillips (1988, 1989), etc. Compared to Lin et al. (2020), the current model incorporates time trends, multivariate I(1) processes and stationary regressors, though these components are linearly related with the monotonic transformation of the dependent variable.

Transformation models have been important tools to analyze economic and financial data. Since Box and Cox (1964) and Bickel and Doksum (1981), a large body of literature has been developed. See, for example, Han (1987) and Abrevaya (1999) for rank estimation of the transformation model; Breiman and Friedman (1985) and Wang and Ruppert (1995) for transform-both-sides models; Chen (2002) and Horowitz (1996) for $\sqrt{n}$-consistent semiparametric estimators of a regression model with an unknown transformation of the dependent variable; Fan and Fine (2013) for linear transformation models with parametric covariate transformation; Chiappori et al. (2015) for identification and estimation of nonparametric transformations; Lewbel et al. (2015) for a specification test for nonparametric transformation models. More recently, Florens and Sokullu (2017), Vanhems and Van Keilegom (2019), and Lin and Tu (2021a) studied semiparametric transformation models in the presence of endogeneity. For more references on this literature, see Lin and Tu (2021a), Lin et al. (2020) and references therein.

This paper first presents an estimation strategy for the proposed model. An extremum estimator of the transformation parameter is proposed via the loss function that measures the relative variation of the regression residual compared to the variation in the transformed dependent variable (Breiman and Friedman, 1985; Lin et al., 2020). The plug-in estimator for rest parameters in the linear component is then obtained. Second, asymptotic distributions for the extremum estimator and the plug-in estimators are then established under a set of regularity conditions. In particular, the limiting distribution of the transformation parameter estimator is nonstandard, with the rate of convergence depending on the model parameters, the properties of the transformation and time trend. For unit root and time trend regressors, the slope estimators converge at order n and $\sqrt{n}\boldsymbol{\kappa}_{nd}$ (the order of time trend), respectively, and have nonstandard distributions that involve functionals of Brownian motions. The estimators for the slope parameters before stationary regressors are shown to be $\sqrt{n}$-consistent and asymptotically normal. The derivations build upon Park and Phillips (2001), Chan and Wang (2015) and Hu et al. (2021), which considered nonlinear parametric regressions with univariate I(1) regressor, and Chang et al. (2001), which studied a nonlinear additive parametric model that accommodates all three types of regressors as in the proposed model. Finally, numerical studies illustrate the merit of our proposed estimators. Simulation results show that the biases of the proposed estimators are small, and their variances decay to zero as the sample size increases. The sampling behavior of the t-ratios associated with the estimators largely corroborates with our theoretical results, and this finding is robust to various choices of parameters.

The rest of this paper is organized as follows. Section 2 introduces the model and the estimators, whose asymptotic properties are presented in Section 3. Section 4 reports some simulation results. Section 5 concludes the paper. The proof of the main theorem is contained in the Appendix, while additional technical details and simulation results are relegated to the online Supplementary Document.

Notations. Throughout the paper, convergence in probability and convergence in distribution are denoted as $\xrightarrow{p}$ and $\Rightarrow$, respectively; and $\boldsymbol{A}^\top$ refers to the transpose of the matrix $\boldsymbol{A}$.

2 Model and estimation

The transformation model of interest is given by

$$\begin{aligned}\Lambda(y_t, \boldsymbol{\beta}_0) &= \boldsymbol{w}_t^\top \boldsymbol{\theta}_0 + u_t \\ &= \boldsymbol{x}_t^\top \boldsymbol{\theta}_1^0 + \boldsymbol{z}_t^\top \boldsymbol{\theta}_2^0 + \boldsymbol{d}_t^\top \boldsymbol{\theta}_3^0 + u_t, \end{aligned} \tag{2.1}$$

for $t = 1, 2, \cdots, n$, where $\Lambda : \mathbb{R} \times \mathbb{R}^{\ell_0} \to \mathbb{R}$ is a known strictly increasing function, $\boldsymbol{x}_t$ is an ℓ_1-dimensional integrated process of order one, and $\boldsymbol{z}_t$ is an ℓ_2-dimensional stationary variable, $\boldsymbol{d}_t$ is an ℓ_3-dimensional deterministic sequence, u_t is the stationary error, and $\boldsymbol{\vartheta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\theta}_1^{0\top}, \boldsymbol{\theta}_2^{0\top}, \boldsymbol{\theta}_3^{0\top})^\top$ is the unknown true parameter vector. This model extends the transformation model of Han (1987) to the cases with time trend and nonstationary regressors. When Λ is specialized to an identity function, (2.1) reduces to the conventional linear cointegrating regression, possibly with a time trend and other stationary regressors, which has been developed in the earlier work of Phillips and Durlauf (1986) and Park and Phillips (1988, 1989). Compared to Lin et al. (2020), this model can allow for multivariate unit root regressors as in Lin and Tu (2021b), but also accommodates a time trend component, which leads to much more complication in the resulting limiting theory. In the specification (2.1), the integrated processes, the deterministic and stationary regressors are assumed to be additively separable. The assumption of additive separability here is not essential, but significantly simplifies the subsequent theoretical development.

The estimators of the unknown parameters are defined sequentially. First, for given $\boldsymbol{\beta}$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \boldsymbol{\theta}_3^\top)^\top$ is estimated by the least squares method, that is,

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}) = (\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{\Lambda}(\boldsymbol{\beta}), \tag{2.2}$$

where $\boldsymbol{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)^\top$, $\boldsymbol{w}_t = (\boldsymbol{x}_t^\top, \boldsymbol{z}_t^\top, \boldsymbol{d}_t^\top)^\top$, $\boldsymbol{\Lambda}(\boldsymbol{\beta}) = (\Lambda(y_1, \boldsymbol{\beta}), \cdots, \Lambda(y_n, \boldsymbol{\beta}))^\top$. Second, for fixed $\boldsymbol{\beta}$, define the loss function

$$L_n(\boldsymbol{\beta}) = \frac{\sum_{t=1}^n [\Lambda(y_t, \boldsymbol{\beta}) - \boldsymbol{w}_t^\top \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})]^2}{\sum_{t=1}^n \Lambda(y_t, \boldsymbol{\beta})^2}. \tag{2.3}$$

Then, the extremum estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0$ is obtained by minimizing $L_n(\boldsymbol{\beta})$ over $\boldsymbol{\beta} \in \Theta_0$, that is, $\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta} \in \Theta_0} L_n(\boldsymbol{\beta})$. Consequently, a plug-in estimator for $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_1^{0\top}, \boldsymbol{\theta}_2^{0\top}, \boldsymbol{\theta}_3^{0\top})^\top$ is defined as $\widehat{\boldsymbol{\theta}}_n = (\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{\Lambda}(\widehat{\boldsymbol{\beta}}_n)$.

The loss function in (2.3) has a normalizing denominator, which is different from that used for the standard nonlinear regressions (e.g., Park and Phillips (2001), and Chan and Wang (2015)). Since there are unknown parameters on the both sides of (2.1), the direct least squares estimation (i.e., without the normalizing denominator) for (2.1) tends to choose $\boldsymbol{\vartheta}$ such that Λ has little variation. On the other hand, minimizing the loss function (2.3) is equivalent to minimizing the fraction of variance in $\Lambda(y, \boldsymbol{\beta})$ not explained by $\boldsymbol{w}_t^\top \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})$. Furthermore, the normalization excludes the trivial specification $\Lambda(y, \beta) = \beta y$ or y/β, under which the loss function in (2.3) is invariant to β. See Breiman and Friedman (1985), Lin et al. (2020), Lin and Tu (2021b, 2020) for a similar objective function used in estimating transformation models and related discussions.

3 Asymptotic theory

3.1 Assumptions

To present the main theorem, the following assumptions are needed.

Assumption 1 *(a) Define* $\boldsymbol{\zeta}_t = (u_t, \boldsymbol{\varepsilon}'_{t+1}, \boldsymbol{\eta}'_{t+1})'$ *and the filtration* $\mathcal{F}_{nt} = \sigma\{(\boldsymbol{\zeta}_s)_{-\infty}^t\}$*, i.e., the* σ*-field generated by* $(\boldsymbol{\zeta}_s)_{s\le t}$*.* $\{\boldsymbol{\zeta}_t, \mathcal{F}_{nt}\}$ *is a stationary and ergodic martingale difference sequence with* $E(\boldsymbol{\zeta}_t\boldsymbol{\zeta}'_t|\mathcal{F}_{n,t-1}) = \boldsymbol{\Sigma}$ *and* $\sup_{t\ge1} E(\|\boldsymbol{\zeta}_t\|^r|\mathcal{F}_{n,t-1}) < \infty$ *for some* $r > 4$*.*

(b) Let $\Delta\boldsymbol{x}_t = \boldsymbol{v}_t = \boldsymbol{\varphi}(L)\boldsymbol{\varepsilon}_t = \sum_{i=0}^\infty \boldsymbol{\varphi}_i\boldsymbol{\varepsilon}_{t-i}$ *and* $\boldsymbol{z}_t = \boldsymbol{\phi}(L)\boldsymbol{\eta}_t = \sum_{i=0}^\infty \boldsymbol{\phi}_i\boldsymbol{\eta}_{t-i}$*, with* $\boldsymbol{\varphi}_0 = \boldsymbol{I}_{\ell_1}$*,* $\boldsymbol{\phi}_0 = \boldsymbol{I}_{\ell_2}$*. Furthermore,* $\boldsymbol{\varphi}(1) = \sum_{i=0}^\infty \boldsymbol{\varphi}_i$ *is nonsingular,* $\sum_{k=0}^\infty k\|\boldsymbol{\varphi}_k\| < \infty$*, and* $\sum_{k=0}^\infty k^{1/2}\|\boldsymbol{\phi}_k\| < \infty$*.*

(c) There exists a nonsingular sequence of normalizing matrices $\boldsymbol{\kappa}_{nd}$ *such that if* $\boldsymbol{d}_n(r) = \boldsymbol{\kappa}_{nd}^{-1}\boldsymbol{d}_{[nr]}$ *on* $[0,1]$*, then* $\sup_{n\ge1}\sup_{0\le r\le1}\|\boldsymbol{d}_n(r)\| \le \infty$*, and* $\boldsymbol{d}_n \xrightarrow{L^2} \boldsymbol{d}$ *for some* $\boldsymbol{d} \in L^2[0,1]$ *such that* $\int_0^1 \boldsymbol{d}(r)\boldsymbol{d}(r)^\top dr > 0$*. The order of the first component of* $\boldsymbol{d}_t$*, denoted as* κ_{nd1}*, is largest among the components of* $\boldsymbol{\kappa}_{nd}$*.*

Assumption 1 (a) and (b) stipulate that the regressor $\boldsymbol{x}_t$ is an integrated process generated by a linear process $\boldsymbol{v}_t$, which has the martingale difference sequence $\{\boldsymbol{\varepsilon}_j, -\infty < j < \infty\}$ as building blocks, $\boldsymbol{z}_t$ is stationary, ergodic and could be correlated with $\boldsymbol{x}_t$. In addition, the regressors $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ are predetermined, i.e., $E(\boldsymbol{x}_t|\mathcal{F}_{n,t-1}) = \boldsymbol{x}_t$ and $E(\boldsymbol{z}_t|\mathcal{F}_{n,t-1}) = \boldsymbol{z}_t$. The nonsingularity of $\boldsymbol{\varphi}(1)$ implies that there is no cointegrating relationship among the component time series in $\boldsymbol{x}_t$. See, for example, Phillips and Solo (1992) for more discussions on these conditions. The conditions in (c) are general enough to allow for deterministic regressors such as constant and time polynomials, possibly with breaks, which are commonly used in time series analyses (see Park (1992), for the asymptotics of integrated processes with such time trends). The convergence here is quite weak, as in most cases of practical interest we will have uniform convergence $\|\boldsymbol{d}_n - \boldsymbol{d}\| \to 0$.

For u_t and $\boldsymbol{v}_t$, we define stochastic processes

$$U_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} u_t, \qquad \text{and} \qquad \boldsymbol{V}_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} \boldsymbol{v}_t, \tag{3.1}$$

on $[0,1]$, where $\lfloor s\rfloor$ denotes the largest integer not exceeding s. The process $(U_n, \boldsymbol{V}_n^\top)^\top$ is defined in $D[0,1]^{1+\ell_1}$, where $D[0,1]$ is the space of cadlag functions on $[0,1]$. Under Assumption 1, an invariance principle holds for $(U_n, \boldsymbol{V}_n^\top)^\top$. That is, we have as $n \to \infty$, $(U_n, \boldsymbol{V}_n^\top)^\top \Rightarrow (U, \boldsymbol{V}^\top)^\top$, where $(U, \boldsymbol{V}^\top)^\top$ is $(1+\ell_1)$-dimensional vector Brownian motion, as shown in Phillips and Solo (1992).

The martingale difference assumption on the regression errors in Assumption 1 (a) is standard in nonlinear nonstationary time series regression, such as Park and Phillips (2000, 2001), Chang et al. (2001), Dong et al. (2016) among others, and it can be relaxed like Chang and Park (2011), Chan and Wang (2015) and Hu et al. (2021). However, a generalization of our theory allowing for correlated errors would involve substantial level of complexity and is not attempted here. However, it does not seem overly restrictive at this point to assume the absence of serial correlation in the errors, especially given our flexible nonlinear specification of the transformation function and the inclusion of stationary regressors in the model.

The following assumption is standard in the extremum estimation theory.

Assumption 2 $\Theta = \Theta_0 \times \Theta_1 \times \Theta_2 \times \Theta_3$, *where* $\Theta_0 \subset \mathbb{R}^{\ell_0}$, $\Theta_1 \subset \mathbb{R}^{\ell_1}$, $\Theta_2 \subset \mathbb{R}^{\ell_2}$ *and* $\Theta_3 \subset \mathbb{R}^{\ell_3}$ *are convex and compact. And the* ℓ*-dimensional true parameter vector* $\boldsymbol{\vartheta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\theta}_1^{0\top}, \boldsymbol{\theta}_2^{0\top}, \boldsymbol{\theta}_3^{0\top})^\top$ *is an interior point of* Θ.

For the ease of presentation, we define $\boldsymbol{\xi}(x,\boldsymbol{\beta}) = \dot{\boldsymbol{\Lambda}}(\Lambda^{-1}(x,\boldsymbol{\beta}_0),\boldsymbol{\beta})$, $\dot{\boldsymbol{\xi}}(x,\boldsymbol{\beta}) = \ddot{\boldsymbol{\Lambda}}(\Lambda^{-1}(x,\boldsymbol{\beta}_0),\boldsymbol{\beta})$, $\ddot{\boldsymbol{\xi}}(x,\boldsymbol{\beta}) = \dddot{\boldsymbol{\Lambda}}(\Lambda^{-1}(x,\boldsymbol{\beta}_0),\boldsymbol{\beta})$, where $\Lambda^{-1}(x,\boldsymbol{\beta})$ is the inverse of $\Lambda(x,\boldsymbol{\beta})$ with respect to x, $\dot{\boldsymbol{\Lambda}}(x,\boldsymbol{\beta}) = \left(\partial\Lambda(x,\boldsymbol{\beta})/\partial\beta_i\right)_{\ell_0\times 1}$, $\ddot{\boldsymbol{\Lambda}}(x,\boldsymbol{\beta}) = \left(\partial^2\Lambda(x,\boldsymbol{\beta})/\partial\beta_i\partial\beta_j\right)_{\ell_0^2\times 1}$, $\dddot{\boldsymbol{\Lambda}}(x,\boldsymbol{\beta}) = \left(\partial^3\Lambda(x,\boldsymbol{\beta})/\partial\beta_i\partial\beta_j\partial\beta_k\right)_{\ell_0^3\times 1}$ are vectors, arrange by the lexicographic ordering of their indices.

Assumption 3 *(a)* $\Lambda(\cdot,\boldsymbol{\beta})$ *is strictly increasing for any given* $\boldsymbol{\beta}$ *and is supposed to be three times continuously differentiable with respect to* $\boldsymbol{\beta}$. $\boldsymbol{\xi}(\cdot,\boldsymbol{\beta})$ *is an H-regular function with asymptotic order* $\boldsymbol{\kappa}_{\boldsymbol{\xi}}(\cdot,\boldsymbol{\beta})$*, and limiting homogeneous function* $\boldsymbol{h}_{\boldsymbol{\xi}}(\cdot,\boldsymbol{\beta})$.

(b) Define a neighborhood of $\boldsymbol{\beta}_0$ *by* $N(\varepsilon,\omega_1,\omega_2) = \{\boldsymbol{\beta} : \|\boldsymbol{\kappa}_{\boldsymbol{\xi}}(\omega_1,\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \le \omega_2^{-1+\varepsilon}\}$ *for given* $\varepsilon > 0$*. For any given* $\bar{s} > 0$*, there exists* $\varepsilon > 0$ *such that as* $\omega_1,\omega_2 \to \infty$,

$$\left\|(\boldsymbol{\kappa}_{\boldsymbol{\xi}} \otimes \boldsymbol{\kappa}_{\boldsymbol{\xi}})^{-1}(\omega_1,\boldsymbol{\beta}_0)\Big(\sup_{|s|\le\bar{s}} \dot{\boldsymbol{\xi}}(\omega_1 s,\boldsymbol{\beta}_0)\Big)\right\| \to 0, \tag{3.2}$$

$$\omega_2^{-1+\varepsilon}\left\|\boldsymbol{\kappa}_{\boldsymbol{\xi}}^{-1}(\omega_1,\boldsymbol{\beta}_0)\Big(\sup_{|s|\le\bar{s}}\sup_{\boldsymbol{\beta}\in N(\varepsilon,\omega_1,\omega_2)} \boldsymbol{\xi}(\omega_1 s,\boldsymbol{\beta})\Big)\right\| \to 0, \tag{3.3}$$

$$\omega_2^{-1+\varepsilon}\left\|(\boldsymbol{\kappa}_{\boldsymbol{\xi}} \otimes \boldsymbol{\kappa}_{\boldsymbol{\xi}})^{-1}(\omega_1,\boldsymbol{\beta}_0)\Big(\sup_{|s|\le\bar{s}}\sup_{\boldsymbol{\beta}\in N(\varepsilon,\omega_1,\omega_2)} \dot{\boldsymbol{\xi}}(\omega_1 s,\boldsymbol{\beta})\Big)\right\| \to 0, \tag{3.4}$$

$$\omega_2^{-1+\varepsilon}\left\|(\boldsymbol{\kappa}_{\boldsymbol{\xi}} \otimes \boldsymbol{\kappa}_{\boldsymbol{\xi}} \otimes \boldsymbol{\kappa}_{\boldsymbol{\xi}})^{-1}(\omega_1,\boldsymbol{\beta}_0)\Big(\sup_{|s|\le\bar{s}}\sup_{\boldsymbol{\beta}\in N(\varepsilon,\omega_1,\omega_2)} \ddot{\boldsymbol{\xi}}(\omega_1 s,\boldsymbol{\beta})\Big)\right\| \to 0, \tag{3.5}$$

where $\boldsymbol{\kappa}_{\boldsymbol{\xi}}$ *are defined in Assumption 3 (a).*

The strictly increasing property of Λ in Assumption 3 (a) is commonly imposed for identification in transformation models. Assumption 3 (a) further stipulates that the function $\boldsymbol{\xi}$ is H-regular (see Definition B.2 in the Supplementary Document), and $\boldsymbol{\kappa}_{\boldsymbol{\xi}}$,

$\boldsymbol{h}_{\boldsymbol{\xi}}$ may depend on $\boldsymbol{\beta}_0$. For theoretical derivation, we can also consider the integrable function class as in Park and Phillips (2001), and the corresponding limiting theory can be obtained following Park and Phillips (2000), Chang and Park (2003) and Dong et al. (2016), etc. However, it is very hard to find a transformation function Λ such that the corresponding composite function $\boldsymbol{\xi}$ is integrable. Thus, we do not consider the integrable function class from the practical point of view and leave it as a future work. Assumption 3 (b) is similar to Assumption (b) of Theorem 5.3 in Park and Phillips (2001), and is required to prove a uniform convergence result. It holds for many H-regular functions used in nonlinear analysis. In addition, we may replace (3.3)-(3.5) with stronger, yet easier to verify conditions. See Park and Phillips (2001) for more related details.

3.2 Distribution theory

The following gives the asymptotic distributions for the estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\theta}}_n$.

Theorem 1 *Let Assumptions 1-3 hold. Assume that for all $\delta > 0$,*

$$\int_{|s|\le\delta} \boldsymbol{h}_{\boldsymbol{\xi},0}(s)\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(s)ds > 0. \tag{3.6}$$

Then the following assertions hold as $n \to \infty$.

(a) If $\kappa_{nd1}/\sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$,

$$\begin{aligned}\sqrt{n}\boldsymbol{\kappa}_{\boldsymbol{\xi}}(\sqrt{n}, \boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) &\Rightarrow \boldsymbol{\mathcal{B}}_1 \\ &\equiv -\Big[\int_0^1 \tau_3^2(r)dr \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_3(r))dr\Big]^{-1} \\ &\times \Big\{\Big(\int_0^1 \tau_3^2(r)dr\Big)\Big[\int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))dU(r) + \sigma_u^2 \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}^x(\tau_3(r))dr\Big] \\ &- \sigma_u^2\Big[\int_0^1 \tau_1(r)\boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))dr + \pi\int_0^1 \tau_2(r)\boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))dr\Big]\Big\},\end{aligned}$$

and $\boldsymbol{D}_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \Rightarrow \boldsymbol{\Delta}_J^{-1}\boldsymbol{\Delta}_S + \begin{pmatrix}\boldsymbol{\mathcal{A}}_1^\top & \mathbf{0}^\top & \boldsymbol{\mathcal{A}}_2^\top\end{pmatrix}^\top$*;*

(b) if $\sqrt{n}/\kappa_{nd1} \to 0$,

$$\begin{aligned}\sqrt{n}\boldsymbol{\kappa}_{\boldsymbol{\xi}}(\kappa_{nd1}, \boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) &\Rightarrow \boldsymbol{\mathcal{B}}_2 \\ &\equiv -\Big(\int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_2(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_2(r))dr\Big)^{-1}\int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_2(r))dU(r),\end{aligned}$$

and $\boldsymbol{D}_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \Rightarrow \boldsymbol{\Delta}_J^{-1}\boldsymbol{\Delta}_S + \begin{pmatrix}\boldsymbol{\mathcal{A}}_3^\top & \mathbf{0}^\top & \boldsymbol{\mathcal{A}}_4^\top\end{pmatrix}^\top$*,* *where*

$$\boldsymbol{D}_n = \begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0 & 0 \\ 0 & \sqrt{n}\boldsymbol{I}_{\ell_2} & 0 \\ 0 & 0 & \sqrt{n}\boldsymbol{\kappa}_{nd} \end{pmatrix}, \boldsymbol{\Delta}_S = \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)dU(r) \\ N(0, \boldsymbol{\Omega}_{zu}) \\ \int_0^1 \boldsymbol{d}(r)dU(r) \end{pmatrix},$$

$$\boldsymbol{\Delta}_J = \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)\boldsymbol{V}^\top(r)dr & 0 & \int_0^1 \boldsymbol{V}(r)\boldsymbol{d}^\top(r)dr \\ 0 & E[\boldsymbol{z}_t\boldsymbol{z}_t^\top] & 0 \\ \int_0^1 \boldsymbol{d}(r)\boldsymbol{V}^\top(r)dr & 0 & \int_0^1 \boldsymbol{d}(r)\boldsymbol{d}^\top(r)dr \end{pmatrix},$$

$\tau_1(r) = \boldsymbol{V}(r)^\top\boldsymbol{\theta}_1^0$, $\tau_2(r) = d_1(r)\theta_{31}^0$, $\tau_3(r) = \tau_1(r) + \pi\tau_2(r)$, $\sigma_u^2 = var(u_t)$, $\boldsymbol{h}_{\boldsymbol{\xi},0}(x) = \boldsymbol{h}_{\boldsymbol{\xi}}(x, \boldsymbol{\beta}_0)$, $\boldsymbol{h}^x_{\boldsymbol{\xi},0}(x) = \partial\boldsymbol{h}_{\boldsymbol{\xi},0}(x)/\partial x$, $\boldsymbol{\Omega}_{zu} = lim_{n\to\infty}n^{-1}var(\sum_{t=1}^n \boldsymbol{z}_t u_t)$, $\boldsymbol{\mathcal{A}}_1 = -\big(\int_0^1 \boldsymbol{V}(r)\boldsymbol{h}^\top_{\boldsymbol{\xi},0}(\tau_3(r))dr\big)\times \boldsymbol{\mathcal{B}}_1$, $\boldsymbol{\mathcal{A}}_2 = -\big(\int_0^1 \boldsymbol{d}(r)\boldsymbol{h}^\top_{\boldsymbol{\xi},0}(\tau_3(r))dr\big) \times \boldsymbol{\mathcal{B}}_1$, $\boldsymbol{\mathcal{A}}_3 = -\big(\int_0^1 \boldsymbol{V}(r)\boldsymbol{h}^\top_{\boldsymbol{\xi},0}(\tau_2(r))dr\big) \times \boldsymbol{\mathcal{B}}_2$, *and* $\boldsymbol{\mathcal{A}}_4 = -\big(\int_0^1 \boldsymbol{d}(r)\boldsymbol{h}^\top_{\boldsymbol{\xi},0}(\tau_2(r))dr\big) \times \boldsymbol{\mathcal{B}}_2$. *Here,* $d_1(r)$ *and* θ_{31}^0 *are the first elements of* $\boldsymbol{d}(r)$ *and* $\boldsymbol{\theta}_3^0$*, respectively.*

Remark 3.1 *The condition (3.6) is introduced for identification purpose and is similar to that of Park and Phillips (2001, Theorem 5.2) and Uematsu (2019, Theorem 3.1). Detailed discussions of these conditions are given in Park and Phillips (2001). Note that the limiting distributions for nonstationary parameters* ($\boldsymbol{\beta}$, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_3$) *in the two cases are nonstandard and likely to have an asymptotic bias. However, how to construct the bias-corrected estimator in this setup remains a challenging issue, and is beyond the scope of this study. See Chang et al. (2001) and Chan and Wang (2015) for similar issues.*

Remark 3.2 *The rate of convergence and limiting distribution of* $\widehat{\boldsymbol{\beta}}_n$ *are nonstandard and depend on* $\boldsymbol{\beta}_0$, Λ *and the relatively size of* κ_{nd1} *and* $\sqrt{n}$. *When* $\kappa_{nd1}/\sqrt{n} \to 0$, $\widehat{\boldsymbol{\beta}}_n$ *converges at rate* $\sqrt{n}\boldsymbol{\kappa}_\xi(\sqrt{n}, \beta_0)$*; when* $\sqrt{n}/\kappa_{nd1} \to 0$*, the convergence rate becomes* $\sqrt{n}\boldsymbol{\kappa}_\xi(\kappa_{nd1}, \beta_0)$. *The result is similar to that in Lin et al. (2020), which also found that the limiting distribution of the transformation parameter is nonstandard with the rate of convergence depending on the property of the unknown transformations. In addition, the convergence rates of the estimators for the coefficients of the stationary, unit root and deterministic regressors are given respectively by* $\sqrt{n}$, n *and* $\sqrt{n}\boldsymbol{\kappa}_{nd}$*, as in standard regressions. The limiting distribution of the estimators of the coefficients before the stationary regressors is normal. However, the asymptotic distributions of the estimators of time trend or unit root coefficients are generally non-Gaussian. As shown in Theorem 1, the limiting distribution for the estimators of time trend or unit root coefficients contains two parts. The first part depends on* $\boldsymbol{V}(r)$ *and* $U(r)$*, and is Gaussian as long as* $\boldsymbol{V}$ *and* U *are asymptotically independent; the second part accounts for the estimation effect induced by the transformation parameters. Such a finding is brand new in the literature and has not been discovered in Chang et al. (2001), Lin et al. (2020), Lin and Tu (2021b), etc. The limiting distributions are thus not centered, and the usual chi-squared approach to inference is not possible. The* *critical values of the usual tests are dependent upon nuisance parameters. See Park and Phillips (2001) and Chang and Park (2003) etc., for similar discussions. While in some specific cases, these limiting distributions will degenerate to Gaussian. For example, when the transformation Λ is the Box-Cox function, and $\boldsymbol{V}(r)$ and $U(r)$ are asymptotically independent, the limiting distribution for the estimators of the time trend and unit root coefficients is mixed normal and the conventional inference can be applied. See Examples 3.1-3.2 and Section 4 below for details.*

Remark 3.3 *The limiting distribution of $\widehat{\boldsymbol{\theta}}_{2n}$ is not affected by $\widehat{\boldsymbol{\beta}}_n$, and is identical to that of the least squares estimator in $\Lambda(y_t, \boldsymbol{\beta}_0) = \boldsymbol{z}_t^\top \boldsymbol{\theta}_2^0 + u_t$. That is, the stationary regressors are asymptotically orthogonal to the integrated regressors and the deterministic trends, as found in Chang et al. (2001). However, the asymptotic distributions of $\widehat{\boldsymbol{\theta}}_{1n}$ and $\widehat{\boldsymbol{\theta}}_{3n}$ both depend on that of $\widehat{\boldsymbol{\beta}}_n$, thus affect each other. This finding is different from the findings in Chang et al. (2001) and Kim and Kim (2012), because of the presence of transformation function Λ.*

Example 3.1 *As an illustrative example, we look at the regression with the constant deterministic component $d_t = 1$. Consider the Box-Cox transformation (Box and Cox, 1964) function $\Lambda(y, \beta) = (y^\beta - 1)/\beta$, $\beta \neq 0$, it is easy to show that $\xi(y, \beta) = (\beta\beta_0)^{-1}(\beta_0 y + 1)^{\beta/\beta_0} \ln(\beta_0 y + 1) - \beta^{-2}(\beta_0 y + 1)^{\beta/\beta_0} + \beta^{-2}$, the homogeneous function is $h_\xi(y, \beta) = \frac{1}{\beta\beta_0}(\beta_0 y)^{\beta/\beta_0}$, and the asymptotic order is $\kappa_\xi(\lambda, \beta) = \lambda^{\beta/\beta_0} \ln(\lambda)$. Then, by Theorem 1, we have*

$$n \ln(\sqrt{n})(\widehat{\beta}_n/\beta_0 - 1) \Rightarrow \mathcal{B}_1 = -\Big(\int_0^1 \tau_1^2(r)dr\Big)^{-1} \int_0^1 \tau_1(r)dU(r), \tag{3.7}$$

$$\begin{aligned}\begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0 \\ 0 & \sqrt{n} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1^0 \\ \widehat{\theta}_{n3} - \theta_3^0 \end{pmatrix} \Rightarrow{}& \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)\boldsymbol{V}^\top(r)dr & \int_0^1 \boldsymbol{V}(r)dr \\ \int_0^1 \boldsymbol{V}^\top(r)dr & 1 \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)dU(r) \\ \int_0^1 dU(r) \end{pmatrix} \\ &- \begin{pmatrix} \big(\int_0^1 \boldsymbol{V}(r)\tau_1(r)dr\big) \times \mathcal{B}_1 \\ \big(\int_0^1 \tau_1(r)dr\big) \times \mathcal{B}_1 \end{pmatrix},\end{aligned} \tag{3.8}$$

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{n2} - \boldsymbol{\theta}_2^0) \Rightarrow E[\boldsymbol{z}_t\boldsymbol{z}_t^\top]^{-1} N(0, \boldsymbol{\Omega}_{zu}). \tag{3.9}$$

In addition, we consider the monotonic transformation function $\Lambda(y, \beta) = e^{\beta y}$. Similarly, by a simple calculation, we have $\xi = \frac{1}{\beta_0} y^{\beta/\beta_0} \ln(y)$, with the homogeneous function $h_\xi(y, \beta) = \frac{1}{\beta_0} y^{\beta/\beta_0}$ and the asymptotic order $\kappa_\xi(\lambda, \beta) = \lambda^{\beta/\beta_0} \ln(\lambda)$. The limiting results for model

$$e^{\beta y_t} = \boldsymbol{x}_t^\top \boldsymbol{\theta}_1^0 + \boldsymbol{z}_t^\top \boldsymbol{\theta}_2^0 + \boldsymbol{d}_t^\top \boldsymbol{\theta}_3^0 + u_t,$$

can be obtained following Theorem 1, which are the same as (3.7)-(3.9).

Example 3.2 *Consider the model with the linear deterministic component $d_t = t$ and the Box-Cox transformation $\Lambda(y, \beta) = (y^\beta - 1)/\beta$, $\beta \neq 0$. By Theorem 1, we have*

$$n^{3/2}\ln(n)(\widehat{\beta}_n/\beta_0 - 1) \Rightarrow \mathcal{B}_2 = -\Big(\int_0^1 \tau_2^2(r)dr\Big)^{-1}\int_0^1 \tau_2(r)dU(r), \tag{3.10}$$

$$\begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0 \\ 0 & n^{3/2} \end{pmatrix}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1^0 \\ \widehat{\theta}_{n3} - \theta_3^0 \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)\boldsymbol{V}^\top(r)^{-1}dr & \int_0^1 \boldsymbol{V}(r)dr \\ \int_0^1 \boldsymbol{V}^\top(r)dU(r) & 1 \end{pmatrix}^{-1}\begin{pmatrix} \int_0^1 \boldsymbol{V}(r)dU(r) \\ \int_0^1 dU(r) \end{pmatrix}$$
$$- \begin{pmatrix} \big(\int_0^1 \boldsymbol{V}(r)\tau_2(r)dr\big) \times \mathcal{B}_2 \\ \big(\int_0^1 \tau_2(r)dr\big) \times \mathcal{B}_2 \end{pmatrix}, \tag{3.11}$$

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{n2} - \boldsymbol{\theta}_2^0) \Rightarrow E[\boldsymbol{z}_t\boldsymbol{z}_t^\top]^{-1}N(0, \boldsymbol{\Omega}_{zu}). \tag{3.12}$$

For the exponential transformation $\Lambda(y, \beta) = e^{\beta y}$, by Theorem 1, we also have (3.10)-(3.12) hold.

The finite sample performance of the estimators in the above two examples shall be investigated in the following section via Monte Carlo simulations.

4 Simulation

This section investigates the finite sample performance of the proposed estimators. To this end, consider the model

$$\Lambda(y_t, \beta_0) = \boldsymbol{x}_t^\top\boldsymbol{\theta}_1^0 + \boldsymbol{z}_t^\top\boldsymbol{\theta}_2^0 + \theta_3^0 d_t + u_t, \qquad t = 1, 2, \cdots, n,$$

where $d_t \in \{1, t\}$, $\boldsymbol{\theta}_1^0 = (1.5, 1)^\top$, $\boldsymbol{\theta}_2^0 = (0.2, 0.4)^\top$ and $\theta_3^0 = 2$. Let $\boldsymbol{x}_t = (x_{1t}, x_{2t})^\top$ be generated by $\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{v}_t$, where $(\boldsymbol{v}_t^\top, \boldsymbol{z}_t^\top)^\top$ is a bivariate normal vector with zero mean and covariance matrix $\{1, 0, \rho, 0; 0, 1, 0, \rho; \rho, 0, 0.5, 0; 0, \rho, 0, 0.5\}$, for $\rho = 0, 0.3, 0.7$, and the error u_t is $N(0, 0.5^2)$. The transformation function Λ is set as the Box-Cox transformation, $\Lambda(y, \beta_0) = (y^{\beta_0} - 1)/\beta_0$ (M1) and $\Lambda(y, \beta) = e^{\beta y}$ (M2) for $\beta_0 = 0.8, 1, 1.2$. In our simulations, we draw samples of sizes $n = 250, 500$ to obtain the proposed estimators and their t-ratios. The bias, standard deviation (SD) and root of mean squared errors (RMSEs) of each estimator are calculated over $1,000$ replications. Kernel densities of the t-ratios are computed using the standard normal kernel with the rule-of-thumb bandwidth. Due to space limitation, we only report some selected results. Other results are similar and are contained in the Supplementary Document.

The findings are summarized as follows. First, the finite sample performances of the estimators are quite close to what would be expected from the limit theory. As can be seen from Table 1 and 3, the biases of the estimators are close to zero, and the root mean squared errors of the estimators are small and decrease as the sample size increases in all cases. This confirms that our estimators are accurate in all scenarios. The realized ratios of the RMSEs for $n = 250$ to those for $n = 500$ are close to the theoretical counterparts as shown in Table 2 and 4, which are consistent with the convergence rates obtained in Examples 3.1-3.2. In particular, the estimator for transformation parameter $\widehat{\beta}_n$ converges the fastest among all the estimators, for both cases with $d_t = 1$ or $d_t = t$. When $d_t = 1$, the estimators in stationary and time trend parts ($\widehat{\boldsymbol{\theta}}_{2n}$ and $\widehat{\theta}_{3n}$) converge at the same rate, which is slower than that of the estimators for the coefficients before unit root processes $\widehat{\boldsymbol{\theta}}_{1n}$. When $d_t = t$, $\widehat{\theta}_{3n}$ converges faster than $\widehat{\boldsymbol{\theta}}_{1n}$ does, and the latter converges much faster than $\widehat{\boldsymbol{\theta}}_{2n}$ does. Second, the sampling behavior of the t-ratios of the estimators largely corroborates with our theoretical results in Section 3. The asymptotic theory demonstrated in Examples 3.1-3.2 shows that the limiting distributions for all estimators are standard (either normal or mixed normal). It has been verified in Figure 4.1 that the kernel density curves of t-ratios are symmetric and centered around zero, which are approximating the limiting standard normal density reasonably well. These findings are robust to various specifications of Λ, β_0, ρ, d_t and sample size n. Note that in other simulation settings, the limiting distributions for the estimators of the transformation parameters, unit root and time trend coefficients may not centered, and the associated t-ratios will have different performances, which should be discussed case by case.

In addition, we compute the percentages of rejection for t-ratios to test the null hypothesis $H_0 : \varsigma = \varsigma_0 + j$ ($\varsigma = \beta, \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}, \theta_3$), at significance level $\alpha = 0.01, 0.05, 0.10$, where j is taken from a grid formed between -0.20 and 0.20 with an equal increment 0.05. Form the figures in the Supplementary Document, we conclude that the sizes of the t-ratios ($j = 0$) are close to the nominal level α for various combinations of Λ, ρ, β and α. Furthermore, the power of t-ratios ($j \neq 0$) for $\varsigma = \beta, \theta_{11}, \theta_{12}$ is close to 1 when $\varsigma \neq \varsigma_0$ even with sample size 250, and that for θ_{21} and θ_{22} becomes larger as the absolute value of j increases and grows with the sample size n. For θ_3, the performance of power of t-ratios is affected by the choice of d_t. When $d_t = 1$, the power increases as $|j|$ and n become larger; when $d_t = t$, the power is close to 1 for various combinations of n and j. Overall, the t-ratios enjoy nice finite sample performance.

5 Conclusion

This paper considers a transformation model with a time trend, stationary regressors, as well as multiple I(1) regressors, which is a hybrid of a transformation model and a conventional cointegration model. Estimation of the unknown quantities is investigated and an asymptotic theory of the proposed estimators is established. Numerical results demonstrate the nice performance of the estimators and corroborate the limiting results.

Table 1: Bias, SD and RMSE ($\times 10^3$) for $\widehat{\beta}_n$ and $\widehat{\boldsymbol{\theta}}_n$ with $\rho = 0.3$, M1

β_0	n	250						500					
		$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
		$d_t = 1$											
	BIAS	-0.04	-0.12	-0.27	1.43	-1.41	1.33	-0.02	-0.15	0.00	0.71	-1.25	1.58
0.8	SD	0.34	7.41	7.12	45.9	44.1	116	0.14	3.91	3.84	31.4	30.4	89.0
	RMSE	0.34	7.41	7.12	45.9	44.1	116	0.14	3.91	3.84	31.4	30.5	89.0
	BIAS	-0.02	-0.37	0.09	-1.55	-1.34	1.01	-0.01	-0.06	-0.10	-0.84	1.70	1.07
1	SD	0.38	7.24	7.74	45.6	46.4	121	0.17	3.84	3.77	31.7	32.8	87.5
	RMSE	0.38	7.25	7.74	45.7	46.4	121	0.17	3.84	3.77	31.7	32.8	87.5
	BIAS	-0.02	-0.02	-0.15	-1.08	-1.40	-0.87	-0.02	0.09	-0.21	-0.99	-0.01	0.56
1.2	SD	0.41	7.43	7.20	44.4	46.5	126	0.19	3.77	3.95	32.3	32.1	86.9
	RMSE	0.41	7.43	7.20	44.4	46.6	126	0.19	3.77	3.96	32.3	32.1	86.9
		$d_t = t$											
	BIAS	0.00	-0.10	0.08	-0.28	-1.77	0.01	0.00	-0.08	-0.02	-1.39	-1.07	0.00
0.8	SD	0.02	7.17	7.37	47.2	44.8	0.72	0.01	3.74	3.81	31.5	31.5	0.25
	RMSE	0.02	7.17	7.37	47.2	44.8	0.72	0.01	3.74	3.81	31.5	31.5	0.25
	BIAS	0.00	-0.23	-0.09	-0.50	2.06	0.02	0.00	0.02	-0.10	0.27	-1.15	0.00
1	SD	0.02	7.43	7.56	46.1	44.8	0.71	0.01	3.60	3.52	30.1	32.3	0.26
	RMSE	0.02	7.43	7.57	46.1	44.8	0.71	0.01	3.60	3.52	30.1	32.3	0.26
	BIAS	0.00	-0.36	0.20	-2.28	0.76	0.00	0.00	0.05	-0.25	-0.39	0.66	0.01
1.2	SD	0.02	7.00	7.15	44.5	44.8	0.72	0.01	3.72	3.62	32.1	30.5	0.25
	RMSE	0.02	7.01	7.15	44.6	44.8	0.72	0.01	3.72	3.63	32.1	30.5	0.25

Table 2: Ratios of the root mean squared errors for $n = 250$ to those for $n = 500$ with $\rho = 0.3$, M1

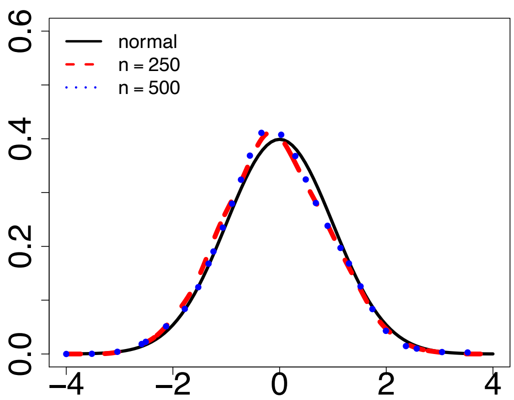
		β_0	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.385	1.895	1.857	1.463	1.449	1.300
		1	2.193	1.885	2.053	1.439	1.413	1.379
		1.2	2.146	1.967	1.818	1.375	1.450	1.450
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	3.138	1.917	1.936	1.499	1.425	2.846
		1	3.093	2.065	2.150	1.533	1.387	2.701
		1.2	3.100	1.882	1.970	1.388	1.470	2.909

Table 3: Bias, SD and RMSE ($\times 10^3$) for $\widehat{\beta}_n$ and $\widehat{\boldsymbol{\theta}}_n$ with $\rho = 0$, M2

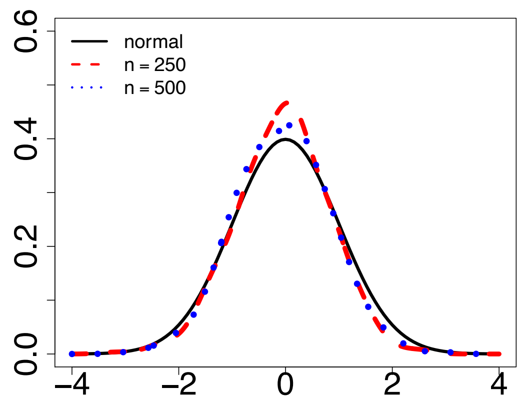
β_0	n	250						500					
		$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
		$d_t = 1$											
	BIAS	-0.01	-0.26	-0.22	2.45	1.03	2.28	0.00	0.21	-0.04	0.29	0.04	-4.96
0.8	SD	0.23	7.41	7.78	44.6	43.0	130	0.11	3.78	3.81	31.9	32.9	91
	RMSE	0.23	7.42	7.78	44.7	43.0	130	0.11	3.78	3.81	31.9	32.9	91
	BIAS	-0.03	-0.27	-0.01	0.63	1.23	-0.24	-0.01	0.00	-0.04	-0.04	0.69	-0.28
1	SD	0.28	6.92	7.00	44.0	46.0	119	0.13	3.68	3.83	32.6	30.9	87
	RMSE	0.28	6.93	7.00	44.0	46.1	119	0.13	3.68	3.83	32.6	30.9	87
	BIAS	-0.02	-0.20	-0.10	-0.76	-1.53	0.65	-0.01	0.00	-0.22	0.59	1.25	3.03
1.2	SD	0.36	7.66	7.54	44.2	46.2	120	0.16	3.90	3.67	32.4	31.6	86
	RMSE	0.36	7.67	7.54	44.3	46.2	120	0.17	3.90	3.67	32.4	31.6	86
		$d_t = t$											
	BIAS	0.00	0.01	0.15	-0.30	-3.49	0.00	0.00	0.00	0.03	1.49	1.57	0.00
0.8	SD	0.01	7.14	7.24	47.1	46.1	0.68	0.00	3.45	3.63	30.8	32.1	0.24
	RMSE	0.01	7.14	7.24	47.1	46.3	0.68	0.00	3.45	3.63	30.9	32.2	0.24
	BIAS	0.00	0.16	-0.31	-1.77	-1.21	0.02	0.00	0.22	-0.09	-0.91	-1.69	-0.01
1	SD	0.02	7.05	7.12	46.3	44.5	0.65	0.01	3.51	3.54	31.6	33.3	0.24
	RMSE	0.02	7.05	7.13	46.4	44.5	0.66	0.01	3.51	3.54	31.6	33.4	0.24
	BIAS	0.00	0.26	-0.20	0.84	-0.45	0.01	0.00	0.06	0.01	-2.81	1.18	-0.01
2	SD	0.02	7.15	7.03	45.4	45.2	0.66	0.01	3.72	3.58	30.7	31.6	0.26
	RMSE	0.02	7.15	7.03	45.4	45.2	0.66	0.01	3.72	3.58	30.9	31.6	0.26

Table 4: Ratios of the root mean squared errors for $n = 250$ to those for $n = 500$ with $\rho = 0$, M2

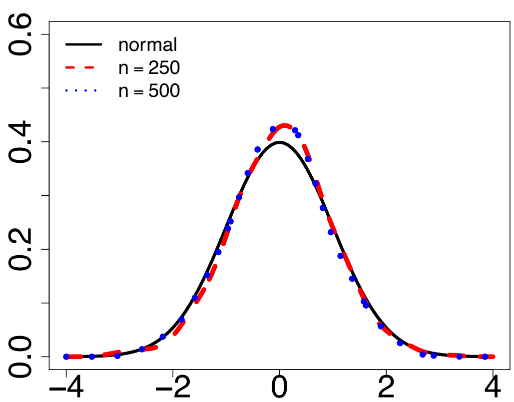
		β_0	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.177	1.961	2.044	1.400	1.310	1.425
		1	2.110	1.881	1.826	1.349	1.491	1.358
		1.2	2.195	1.968	2.052	1.364	1.460	1.391
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	3.135	2.067	1.995	1.527	1.439	2.847
		1	3.095	2.009	2.013	1.468	1.334	2.723
		1.2	2.959	1.920	1.964	1.473	1.429	2.494



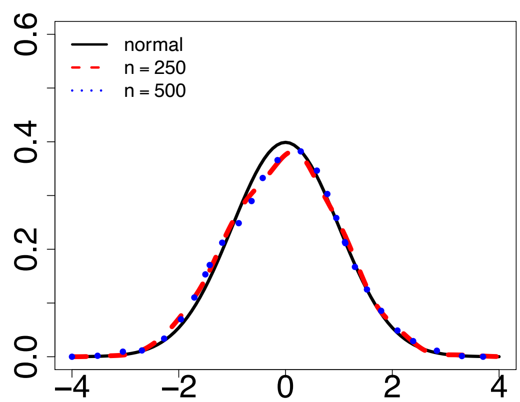
(a) $\widehat{\beta}_n$



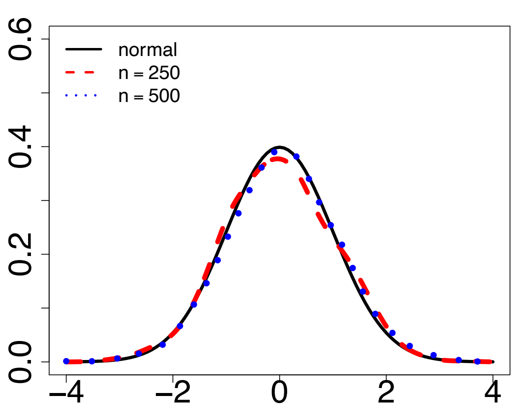
(b) $\widehat{\theta}_{n,11}$



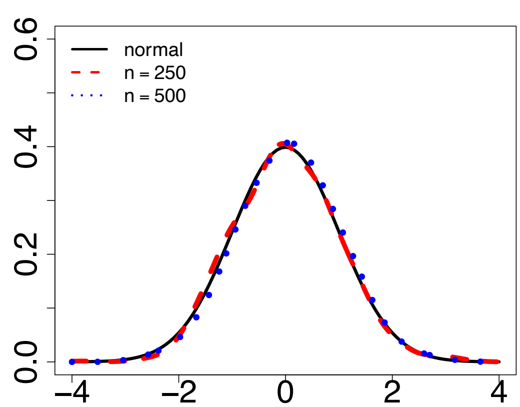
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$



(e) $\widehat{\theta}_{n,22}$



(f) $\widehat{\theta}_{n,3}$

Figure 4.1: Kernel density of t-ratios with $d_t = 1, \beta_0 = 1, \rho = 0$, M1.

There are several possible directions to extend the paper. First, the right-hand side of our model is linear in parameters, which can be extended to the general nonlinear setting, such as the index model of Park and Phillips (2000) and Chang and Park (2003), the additive model of Chang et al. (2001), or the (partially linear) single index model of Dong et al. (2016). Second, endogeneity could be incorporated in the current setting and worths consideration in the future work. Some further development following Chang and Park (2011) and Chan and Wang (2015) could be made. Third, specification test for the parametric form of the transformation and the test for the existence of such cointegration relationship are still underdeveloped. These issues are technically involved and are left as future research.

Acknowledgements

Lin acknowledges the support of the Fundamental Research Funds for the Central Universities (China). Tu (corresponding author) would like to thank support from National Natural Science Foundation of China (Grant 72073002, 12026607, 92046021), the Center for Statistical Science at Peking University, and Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University), Ministry of Education.

Appendix

The proof of Theorem 1 is given below, while all lemmas used in this proof are given in the Supplementary Document.

A Proof of the main theorem

Proof of Theorem 1. We have two cases, i.e., Case (a), if $\kappa_{nd1}/\sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$; Case (b), if $\sqrt{n}/\kappa_{nd1} \to 0$. Because the arguments used to prove the two cases are similar, we present the proofs for Case (a) below, with the detailed proofs for Case (b) omitted. This proof contains four parts. We first investigate the limiting distribution of $\widehat{\boldsymbol{\beta}}_n$. Specifically, Part I gives the score and hessian, Part II establishes their joint asymptotics, and Part III includes a detailed proof for the limit distribution of $\widehat{\boldsymbol{\beta}}_n$. Then Part IV proves that of $\widehat{\boldsymbol{\theta}}_n$.

Part I. The loss function. Since n^2 does not rely on $\boldsymbol{\beta}$, then minimizing $L_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is equivalent to minimizing

$$\widetilde{L}_n(\boldsymbol{\beta}) = \frac{n^2}{2} \frac{\sum_{t=1}^{n} [\Lambda(y_t, \boldsymbol{\beta}) - \boldsymbol{w}_t' \boldsymbol{\theta}_0]^2}{\sum_{t=1}^{n} \Lambda(y_t, \boldsymbol{\beta})^2}. \tag{A.1}$$

Therefore, the score function is

$$S_n(\boldsymbol{\beta}) = n^2 A_{n1}^{-2}(\boldsymbol{\beta})[A_{n1}(\boldsymbol{\beta})A_{n4}(\boldsymbol{\beta}) - A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})], \tag{A.2}$$

and the hessian matrix is

$$\begin{aligned} J_n(\boldsymbol{\beta}) &= n^2 A_{n1}^{-3}(\boldsymbol{\beta})\big\{A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta}) + A_{n6}(\boldsymbol{\beta})] - A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta}) + A_{n7}(\boldsymbol{\beta})] \\ &\quad - 2A_{n1}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})A_{n4}^\top(\boldsymbol{\beta}) + A_{n4}(\boldsymbol{\beta})A_{n3}^\top(\boldsymbol{\beta})] + 4A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})A_{n3}^\top(\boldsymbol{\beta})\big\} \\ &\equiv n^2 J_{n1}(\boldsymbol{\beta})/J_{n2}(\boldsymbol{\beta}), \end{aligned} \tag{A.3}$$

where the definitions of J_{n1}, J_{n2} should be obvious.

Part II. The score and the Hessian. In this case, $\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0} = \boldsymbol{\kappa}_{n\boldsymbol{\xi},0}$. Let $\boldsymbol{D}_n = \sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}$. By Lemma C.2, we have

$$\begin{aligned} \boldsymbol{D}_n^{-1} S_n(\boldsymbol{\beta}_0) &= \big(n^{-2}A_{n1}(\boldsymbol{\beta}_0)\big)^{-2}\big[n^{-2}A_{n1}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n4}(\boldsymbol{\beta}_0) - n^{-1}A_{n2}(\boldsymbol{\beta}_0)(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta}_0)\big] \\ &\Rightarrow \Big(\int_0^1 \tau_3^2(r)dr\Big)^{-2}\Big\{\big(\int_0^1 \tau_3^2(r)dr\big)\Big[\int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))dU(r) + \sigma_u^2\int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}^x(\tau_3(r))dr\Big] \\ &\quad - \sigma_u^2 \int_0^1 \tau_3(r)\boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))dr\Big\}, \\ \boldsymbol{D}_n^{-1} J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top} &= (n^{-2}A_{n1}(\boldsymbol{\beta}_0))^{-1}\cdot(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n5}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1} + o_p(1) \\ &\Rightarrow \left(\int_0^1 \tau_3^2(r)dr\right)^{-1}\int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_3(r))dr. \end{aligned} \tag{A.4}$$

And the joint convergence of S_n and J_n follows from the joint convergence results in Lemma C.2.

Part III. Detailed proof for the limit of $\widehat{\boldsymbol{\beta}}_n$. We shall use Theorem A.1 in Hu et al. (2021) to complete our proof. To apply this theorem, we need show two conditions, that is, (1) the smallest eigenvalue of $\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}$ is positive; and (2)

$$\sup_{\boldsymbol{\beta}\in N_n} \|\boldsymbol{C}_n^{-1}[J_n(\boldsymbol{\beta}) - J_n(\boldsymbol{\beta}_0)]\boldsymbol{C}_n^{-1}\| = o_p(1), \tag{A.5}$$

where $\boldsymbol{C}_n = n^{-\rho}\boldsymbol{D}_n$ for $0 < \rho < \varepsilon/6$ with ε defined in Assumption 3, $N_n = \{\boldsymbol{\beta} : \|\boldsymbol{C}_n(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \le 1\}$.

(1) We first show that $\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}$ is positive definite. Assume that there exists a nonzero vector $\check{c}$ such that $\check{c}^\top\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}\check{c} \le 0$. Then, by *Part II*, we have

$$\begin{aligned} \check{c}^\top\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}\check{c} &= \frac{\check{c}^\top(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n5}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\check{c}}{n^{-2}A_{n1}(\boldsymbol{\beta}_0)} + o_p(1) \\ &\Rightarrow \left(\int_0^1 \tau_3^2(r)dr\right)^{-1}\cdot\check{c}^\top\int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_3(r))dr\check{c} > 0. \end{aligned}$$

Then, $\boldsymbol{D}_n^{-1} J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}$ is positive definite with the smallest eigenvalue larger than 0.

(2) Next, we show (A.5). To do so, first write that

$$
\begin{aligned}
& n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[J_n(\boldsymbol{\beta}) - J_n(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top} \\
& = n^{2\rho+1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\left\{\frac{J_{n2}(\boldsymbol{\beta}) - J_{n2}(\boldsymbol{\beta}_0)}{J_{n1}(\boldsymbol{\beta}_0)} - \frac{J_{n2}(\boldsymbol{\beta}_0)}{J_{n1}^2(\boldsymbol{\beta}_0)}(J_{n1}(\boldsymbol{\beta}) - J_{n1}(\boldsymbol{\beta}_0))\right\}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\frac{J_{n1}(\boldsymbol{\beta}_0)}{J_{n,1}(\boldsymbol{\beta})} \\
& = [n^{-6}J_{n1}(\boldsymbol{\beta}_0)]^{-1}\Big\{n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[J_{n,2}(\boldsymbol{\beta}) - J_{n2}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top} \\
& - (\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}J_{n,11}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-\top}\cdot n^{2\rho-6}[J_{n1}(\boldsymbol{\beta}) - J_{n1}(\boldsymbol{\beta}_0)]\Big\}\cdot\frac{J_{n1}(\boldsymbol{\beta}_0)}{J_{n1}(\boldsymbol{\beta})}.
\end{aligned}
$$

Since by Lemma C.2 (a1), Lemma C.3 (i) and (A.4), we have

$$
\begin{aligned}
& n^{-6}J_{n1}(\boldsymbol{\beta}_0) = O_p(1), \\
& (\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}J_n(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-\top} = O_p(1).
\end{aligned}
$$

To show (A.5), it suffices to show

$$\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[J_{n2}(\boldsymbol{\beta}) - J_{n2}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1), \tag{A.6}$$

$$\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-6}[J_{n1}(\boldsymbol{\beta}) - J_{n1}(\boldsymbol{\beta}_0)]\| = o_p(1), \tag{A.7}$$

$$\sup_{\boldsymbol{\beta}\in N_n}\left\|J_{n1}(\boldsymbol{\beta}_0)/J_{n1}(\boldsymbol{\beta})\right\| = O_p(1). \tag{A.8}$$

Consider (A.6), write

$$
\begin{aligned}
& J_{n2}(\boldsymbol{\beta}) - J_{n2}(\boldsymbol{\beta}_0) \\
& = \left\{A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta}) + A_{n6}(\boldsymbol{\beta})] - A_{n1}^2(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0) + A_{n6}(\boldsymbol{\beta}_0)]\right\} \\
& - \left\{A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta}) + A_{n7}(\boldsymbol{\beta})] - A_{n1}(\boldsymbol{\beta}_0)A_{n2}(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0) + A_{n7}(\boldsymbol{\beta}_0)]\right\} \\
& - 2\Big\{A_{n1}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})A_{n4}^\top(\boldsymbol{\beta}) + A_{n4}(\boldsymbol{\beta})A_{n3}^\top(\boldsymbol{\beta})] - A_{n1}(\boldsymbol{\beta}_0)[A_{n3}(\boldsymbol{\beta}_0)A_{n4}^\top(\boldsymbol{\beta}_0) + A_{n4}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)]\Big\} \\
& + 4\Big\{A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})A_{n3}^\top(\boldsymbol{\beta}) - A_{n2}(\boldsymbol{\beta}_0)A_{n3}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)\Big\} \\
& \equiv \Upsilon_{112}^a - \Upsilon_{112}^b - 2\Upsilon_{112}^c + 4\Upsilon_{112}^d,
\end{aligned}
$$

where the definitions of $\Upsilon_{112}^a - \Upsilon_{112}^d$ should be obvious. For the first term,

$$
\begin{aligned}
\Upsilon_{112}^a = & A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta}) - A_{n5}(\boldsymbol{\beta}_0)] + A_{n1}^2(\boldsymbol{\beta})[A_{n6}(\boldsymbol{\beta}) - A_{n6}(\boldsymbol{\beta}_0)] \\
& + [A_{n1}^2(\boldsymbol{\beta}) - A_{n1}^2(\boldsymbol{\beta}_0)][A_{n5}(\boldsymbol{\beta}_0) + A_{n6}(\boldsymbol{\beta}_0)].
\end{aligned}
$$

By Lemma C.2 and Lemma C.3, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\
&\qquad\le \sup_{\boldsymbol{\beta}\in N_n}\|n^{-4}A_{n1}^2(\boldsymbol{\beta})\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|=o_p(1),\\
&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}^2(\boldsymbol{\beta})[A_{n6}(\boldsymbol{\beta})-A_{n6}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\
&\qquad\le \sup_{\boldsymbol{\beta}\in N_n}\|n^{-4}A_{n1}^2(\boldsymbol{\beta})\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n6}(\boldsymbol{\beta})-A_{n6}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|=o_p(1),\\
&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n1}^2(\boldsymbol{\beta})-A_{n1}^2(\boldsymbol{\beta}_0)][A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\
&\qquad\le \sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-4}[A_{n1}^2(\boldsymbol{\beta})-A_{n1}^2(\boldsymbol{\beta}_0)]\|\cdot\|(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}[A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)](\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-\top}\|\\
&\qquad\le \sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{-2}[A_{n1}(\boldsymbol{\beta})+A_{n1}(\boldsymbol{\beta}_0)]\|\\
&\qquad\quad\times\|(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}[A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)](\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-\top}\|=o_p(1).
\end{aligned}$$

Thus,

$$\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Upsilon_{112}^{a}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|=o_p(1). \tag{A.9}$$

For the second term,

$$\begin{aligned}
\Upsilon_{112}^{b}&=A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta})-A_{n7}(\boldsymbol{\beta}_0)]\\
&\quad+A_{n1}(\boldsymbol{\beta})[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)][A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\\
&\quad+[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]A_{n2}(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)].
\end{aligned}$$

Similarly, by Lemma C.2 and Lemma C.3, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta})-A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|\\
&\qquad\le n^{-1/2}\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{-2}A_{n1}(\boldsymbol{\beta})\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{-1}A_{n2}(\boldsymbol{\beta})\|\\
&\qquad\quad\times\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta})-A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|=o_p(1),\\
&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)][A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|\\
&\qquad= n^{-1/2}\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{-2}A_{n1}(\boldsymbol{\beta})\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-1}[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)]\|\\
&\qquad\quad\times\|n^{-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|=o_p(1),\\
&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]A_{n2}(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|\\
&\qquad\le n^{-1/2}\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]\|\cdot\|n^{-1}A_{n2}(\boldsymbol{\beta}_0)\|\\
&\qquad\quad\times\|n^{-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\|=o_p(1).
\end{aligned}$$

Thus,

$$\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Upsilon_{112}^{b}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1). \tag{A.10}$$

For the third term,

$$\begin{aligned}\Upsilon_{112}^{c} &= A_{n1}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n4}(\boldsymbol{\beta}) - A_{n4}(\boldsymbol{\beta}_0)]^\top + A_{n1}(\boldsymbol{\beta})[A_{n4}(\boldsymbol{\beta}) - A_{n4}(\boldsymbol{\beta}_0)]A_{n3}(\boldsymbol{\beta})^\top \\ &\quad + A_{n1}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]A_{n4}^\top(\boldsymbol{\beta}_0) + A_{n1}(\boldsymbol{\beta})A_{n4}(\boldsymbol{\beta}_0)[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]^\top \\ &\quad + [A_{n1}(\boldsymbol{\beta}) - A_{n1}(\boldsymbol{\beta}_0)][A_{n3}(\boldsymbol{\beta}_0)A_{n4}^\top(\boldsymbol{\beta}_0) + A_{n4}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)].\end{aligned}$$

By Lemma C.2 and Lemma C.3, we have

$$\begin{aligned}&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n4}(\boldsymbol{\beta}) - A_{n4}(\boldsymbol{\beta}_0)]^\top\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\ &\qquad \le n^{-1/2}\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{-2}A_{n1}(\boldsymbol{\beta})\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\|(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta})\| \\ &\qquad\quad \times\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n4}(\boldsymbol{\beta}) - A_{n4}(\boldsymbol{\beta}_0)]\| = o_p(1), \\ &\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})A_{n4}(\boldsymbol{\beta}_0)[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]^\top\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\ &\qquad \le n^{-1}\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{-2}A_{n1}(\boldsymbol{\beta})\|\cdot\|(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n4}(\boldsymbol{\beta}_0)\| \\ &\qquad\quad \times\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]\| = o_p(1), \\ &\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n1}(\boldsymbol{\beta}) - A_{n1}(\boldsymbol{\beta}_0)][A_{n3}(\boldsymbol{\beta}_0)A_{n4}^\top(\boldsymbol{\beta}_0) + A_{n4}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\ &\qquad \le n^{-1}\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta}) - A_{n1}(\boldsymbol{\beta}_0)]\| \\ &\qquad\quad \times\|(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta}_0)[(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n4}(\boldsymbol{\beta}_0)]^\top \\ &\qquad\quad + (\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n4}(\boldsymbol{\beta}_0)[(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta}_0)]^\top\| = o_p(1).\end{aligned}$$

Thus,

$$\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Upsilon_{112}^{c}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1). \tag{A.11}$$

For the last term,

$$\begin{aligned}\Upsilon_{112}^{d} &= A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]^\top + A_{n2}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]A_{n3}^\top(\boldsymbol{\beta}_0) \\ &\quad + [A_{n2}(\boldsymbol{\beta}) - A_{n2}(\boldsymbol{\beta}_0)]A_{n3}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0).\end{aligned}$$

By Lemma C.2 and Lemma C.3, we have

$$\begin{aligned}&\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]^\top\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\ &\qquad = n^{-1}\cdot\sup_{\boldsymbol{\beta}\in N_n}\|n^{-1}A_{n2}(\boldsymbol{\beta})\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\|(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta})\| \\ &\qquad\quad \times\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]\| = o_p(1),\end{aligned}$$

and

$$\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)]A_{n3}(\boldsymbol{\beta}_0)A_{n3}^{\top}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| \\ = n^{-1}\cdot \sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)]\| \cdot \|(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta}_0)\|^2 = o_p(1).$$

Thus,

$$\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Upsilon_{112}^{d}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1). \tag{A.12}$$

Combining (A.9)-(A.12) gives rise to (A.6).

Next consider (A.7). By Lemma C.2 and Lemma C.3,

$$\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-6}[J_{n1}(\boldsymbol{\beta})-J_{n1}(\boldsymbol{\beta}_0)]\| \\ \leq \sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]\| \cdot \sup_{\boldsymbol{\beta}\in N_n} \|n^{-4}[A_{n1}^2(\boldsymbol{\beta})+A_{n1}(\boldsymbol{\beta})A_{n1}(\boldsymbol{\beta}_0)+A_{n1}^2(\boldsymbol{\beta}_0)]\| \\ = o_p(1),$$

showing (A.7). By Lemma C.3, using arguments similar to Theorem 2.2 in Chan and Wang (2014), (A.8) is established.

We have shown in *Part II* the convergence of $S_n(\boldsymbol{\beta}_0)$ and $J_n(\boldsymbol{\beta}_0)$. Thus, by Theorem A.1 in Hu et al. (2021), there exists a sequence of estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0$ such that $S_n(\widehat{\boldsymbol{\beta}}_n)=0$ and the limiting distribution follows.

Part IV. Proof for the limit of $\widehat{\boldsymbol{\theta}}_n$. Finally, we show the limiting distribution of $\widehat{\boldsymbol{\theta}}_n$. Define $\widetilde{\boldsymbol{D}}_n = \begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0 & 0 \\ 0 & \sqrt{n}\boldsymbol{I}_{\ell_2} & 0 \\ 0 & 0 & \sqrt{n}\boldsymbol{\kappa_{nd}} \end{pmatrix}$. Note that

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_n &= (\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{\top}\boldsymbol{\Lambda}(\widehat{\boldsymbol{\beta}}_n) \\ &= (\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{\top}\boldsymbol{\Lambda}(\boldsymbol{\beta}_0) + (\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{\top}[\boldsymbol{\Lambda}(\widehat{\boldsymbol{\beta}}_n)-\boldsymbol{\Lambda}(\boldsymbol{\beta}_0)] \\ &= \boldsymbol{\theta}_0 + (\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{T}u + (\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{\top}[\boldsymbol{\Lambda}(\widehat{\boldsymbol{\beta}}_n)-\boldsymbol{\Lambda}(\boldsymbol{\beta}_0)], \\ \widetilde{\boldsymbol{D}}_n(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0) &= \widetilde{\boldsymbol{D}}_n(\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{T}u + \widetilde{\boldsymbol{D}}_n(\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{\top}[\boldsymbol{\Lambda}(\widehat{\boldsymbol{\beta}}_n)-\boldsymbol{\Lambda}(\boldsymbol{\beta}_0)]. \end{aligned}$$

First, we have

$$\widetilde{\boldsymbol{D}}_n^{-1}\boldsymbol{W}^{\top}\boldsymbol{W}\widetilde{\boldsymbol{D}}_n^{-1} \Rightarrow \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)\boldsymbol{V}'(r)dr & 0 & \int_0^1 \boldsymbol{V}(r)\boldsymbol{d}'(r)dr \\ 0 & E[\boldsymbol{z}_t\boldsymbol{z}_t'] & 0 \\ \int_0^1 \boldsymbol{d}(r)\boldsymbol{V}(r)'dr & 0 & \int_0^1 \boldsymbol{d}(r)\boldsymbol{d}(r)'dr \end{pmatrix} \equiv \boldsymbol{\Delta}_V,$$

by Lemma 5 in Chang et al. (2001). And

$$\widetilde{\boldsymbol{D}}_n(\boldsymbol{W}^{\top}\boldsymbol{W})^{-1}\boldsymbol{W}^{T}u = \boldsymbol{\Delta}_V^{-1}(1+o_p(1))\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t u_t \\ \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t u_t \\ \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t u_t \end{pmatrix} \Rightarrow \boldsymbol{\Delta}_V^{-1}\begin{pmatrix} \int_0^1 \boldsymbol{V}(r)dU(r) \\ N(0,\boldsymbol{\Omega}_{zu}) \\ \int_0^1 \boldsymbol{d}(r)dU(r) \end{pmatrix},$$

where $\boldsymbol{\Omega}_{zu} = lim_{n\to\infty} n^{-1} var(\sum_{t=1}^n \boldsymbol{z}_t u_t)$. In addition, using the mean value theorem, we have

$$\widetilde{\boldsymbol{D}}_n(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top[\boldsymbol{\Lambda}(\widehat{\boldsymbol{\beta}}_n) - \Lambda(\boldsymbol{\beta}_0)] = \boldsymbol{\Delta}_V^{-1}(1+o_p(1))\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t[\Lambda_t(\widehat{\boldsymbol{\beta}}_n) - \Lambda_t(\boldsymbol{\beta}_0)] \\ \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t[\Lambda_t(\widehat{\boldsymbol{\beta}}_n) - \Lambda_t(\boldsymbol{\beta}_0)] \\ \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t[\Lambda_t(\widehat{\boldsymbol{\beta}}_n) - \Lambda_t(\boldsymbol{\beta}_0)] \end{pmatrix}$$

$$= \boldsymbol{\Delta}_V^{-1}(1+o_p(1))\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\ \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \end{pmatrix}.$$

Using arguments similar to Lemma C.2, tt follows from Lemma C.2 and Theorem 1 that

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
&\quad = \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_n) - \dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)]^\top(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
&\quad = \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(1) = \frac{1}{n^{3/2}}\sum_{t=1}^n \boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\cdot\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(1) \\
&\quad \Rightarrow -\int_0^1 \boldsymbol{V}(r)\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_3(r))dr\cdot\boldsymbol{\mathcal{B}}_1, \\
&\frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t[\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n) - \dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)](\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(1) = \frac{1}{n}\sum_{t=1}^n \boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\cdot\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(1) \xrightarrow{p} 0, \\
&\frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
&\quad = \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t[\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n) - \dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)](\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
&\quad = \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(1) = \frac{\boldsymbol{\kappa}_{nd}^{-1}}{n}\sum_{t=1}^n \boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\cdot\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(1) \\
&\quad \Rightarrow -\int_0^1 \boldsymbol{d}(r)\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_3(r))dr\cdot\boldsymbol{\mathcal{B}}_1,
\end{aligned}$$

where $\boldsymbol{\mathcal{B}}_1$ is defined in Theorem 1. Combining the above pieces, with an application of Lemma C.2, completes the proof.

References

Abrevaya, J. (1999). Rank estimation of a transformation model with observed truncation. *Econometrics Journal*, 2(2):292–305.

Bae, Y. and De Jong, R. M. (2007). Money demand function estimation by nonlinear cointegration. *Journal of Applied Econometrics*, 22(4):767–793.

Bae, Y., Kakkar, V., and Ogaki, M. (2006). Money demand in japan and nonlinear cointegration. *Journal of Money, Credit, and Banking*, 38(6):1659–1667.

Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, 76(374):296–311.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–243.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, 80(391):580–598.

Cai, Z., Li, Q., and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113.

Chan, N. and Wang, Q. (2014). Uniform convergence for nonparametric estimators with nonstationary data. *Econometric Theory*, 30(5):1110–1133.

Chan, N. and Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, 185(1):182–195.

Chang, Y., Jiang, B., and Park, J. Y. (2012). Nonstationary regression with logistic transition. *The Econometrics Journal*, 15(2):255–287.

Chang, Y. and Park, J. Y. (2003). Index models with integrated time series. *Journal of Econometrics*, 114(1):73–106.

Chang, Y. and Park, J. Y. (2011). Endogeneity in nonlinear regressions with integrated time series. *Econometric Reviews*, 30(1):51–87.

Chang, Y., Park, J. Y., and Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal*, 4(1):1–36.

Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, 70(4):1683–1697.

Chiappori, P.-A., Komunjer, I., and Kristensen, D. (2015). Nonparametric identification and estimation of transformation models. *Journal of Econometrics*, 188(1):22–39.

Dong, C. and Gao, J. (2018). Specification testing in structural nonparametric cointegration. *Econometric Theory*, 34(4):754–789.

Dong, C., Gao, J., and Tjøstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. *Annals of Statistics*, 44(1):425–453.

Dong, C., Gao, J., Tjøstheim, D., and Yin, J. (2017). Specification testing for nonlinear multivariate cointegrating regressions. *Journal of Econometrics*, 200(1):104–117.

Dong, C. and Linton, O. (2018). Additive nonparametric models with time variable and both stationary and nonstationary regressors. *Journal of Econometrics*, 207(1):212–236.

Dong, C., Linton, O., and Peng, B. (2021). A weighted sieve estimator for nonparametric time series models with nonstationary variables. *Journal of Econometrics*, 222(2):909–932.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Fan, C. and Fine, J. P. (2013). Linear transformation model with parametric covariate transformations. *Journal of the American Statistical Association*, 108(502):701–712.

Florens, J.-P. and Sokullu, S. (2017). Nonparametric estimation of semiparametric transformation models. *Econometric Theory*, 33(4):839–873.

Gao, J. and Phillips, P. C. B. (2013). Functional coefficient nonstationary regression with non- and semi-parametric cointegration. *Working Paper*, Monash University.

Granger, C. (1991). Some recent generalisations of cointegration and the analysis of longrun relationships. In Engle, R. and Granger, C., editors, *Long-Run Economic Relationships*, pages 277–287. Oxford University Press.

Granger, C. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships.* Oxford University Press, New York.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16(1):121–130.

Han, A. K. (1987). A non-parametric analysis of transformations. *Journal of Econometrics*, 35(2-3):191–209.

Hirukawa, M. and Sakudo, M. (2018). Functional-coefficient cointegration models in the presence of deterministic trends. *Econometric Reviews*, 37(5):507–533.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, 64(1):103–137.

Hu, Z., Phillips, P. C. B., and Wang, Q. (2021). Nonlinear cointegrating power function regression with endogeneity. *Econometric Theory*, doi:10.1017/S0266466620000560.

Juhl, T. (2005). Functional-coefficient models under unit root behaviour. *Econometrics Journal*, 8(2):197–213.

Karlsen, H. A., Myklebust, T., and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *The Annals of Statistics*, 35(1):252–299.

Kasparis, I., Andreou, E., and Phillips, P. C. (2015). Nonparametric predictive regression. *Journal of Econometrics*, 185(2):468–494.

Kasparis, I. and Phillips, P. C. B. (2012). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics*, 168(2):270–284.

Kim, C. S. and Kim, I. M. (2012). Partial parametric estimation for nonstationary nonlinear regressions. *Journal of Econometrics*, 167(2):448–457.

Lewbel, A., Lu, X., and Su, L. (2015). Specification testing for transformation models with an application to generalized accelerated failure-time models. *Journal of Econometrics*, 184(1):81–96.

Lin, Y. and Tu, Y. (2020). Sieve extremum estimation of a semiparametric transformation model. *Economics Letters*, 189:109020.

Lin, Y. and Tu, Y. (2021a). Identification and estimation of a semiparametric single index transformation model. *Working Paper*, Peking University.

Lin, Y. and Tu, Y. (2021b). On transformed linear cointegration models. *Economics Letters*, 198:109686.

Lin, Y., Tu, Y., and Yao, Q. (2020). Estimation for double-nonlinear cointegration. *Journal of Econometrics*, 216(1):175–191.

Linton, O. and Wang, Q. (2016). Nonparametric transformation regression with nonstationary data. *Econometric Theory*, 32(1):1–29.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica*, 60(1):119–143.

Park, J. Y. (2006). Nonstationary nonlinearity: An outlook for new opportunities. *In: Corbae, D., Durlauf, S.N., Hansen, B.E. (Eds.), Econometric Theory and Practice*.

Park, J. Y. and Hahn, S. B. (1999). Cointegrating regressions with time varing coefficients. *Econometric Theory*, 15(5):664–703.

Park, J. Y. and Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4(3):468–497.

Park, J. Y. and Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, 5(1):95–131.

Park, J. Y. and Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, 15(3):269–298.

Park, J. Y. and Phillips, P. C. B. (2000). Nonstationary binary choice. *Econometrica*, 68(5):1249–1280.

Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, 69(1):117–161.

Phillips, P. C. B. (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory*, 25(6):1466–1497.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495.

Phillips, P. C. B., Li, D., and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196(1):180–195.

Phillips, P. C. B. and Park, J. Y. (1998). Nonstationary density estimation and kernel autoregression. *Unpublished Manuscript*.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, 20(2):971–1001.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford University Press.

Tu, Y., Liang, H.-Y., and Wang, Q. (2021). Nonparametric inference for quantile cointegrations with stationary covariates. *Journal of Econometrics*, forthcoming.

Tu, Y. and Wang, Y. (2019). Functional coefficient cointegration models subject to time-varying volatility with an application to the purchasing power parity. *Oxford Bulletin of Economics and Statistics*, 81(6):1401–1423.

Tu, Y. and Wang, Y. (2020). Adaptive estimation of heteroskedastic functional-coefficient regressions with an application to fiscal policy evaluation on asset markets. *Econometric Reviews*, 39(3):299–318.

Tu, Y. and Wang, Y. (2021). Spurious functional-coefficient regression models and robust inference with marginal integration. *Journal of Econometrics*, forthcoming.

Uematsu, Y. (2019). Nonstationary nonlinear quantile regression. *Econometric Reviews*, 38(4):386–416.

Vanhems, A. and Van Keilegom, I. (2019). Estimation of a semiparametric transformation model in the presence of endogeneity. *Econometric Theory*, 35(1):73–110.

Wang, N. and Ruppert, D. (1995). Nonparametric estimation of the transformation in the transform-both-sides regression model. *Journal of the American Statistical Association*, 90(430):522–534.

Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression*. World Scientific.

Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25(3):710–738.

Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, 77(6):1901–1948.

Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, 40(2):727–758.

Wang, Q. and Phillips, P. C. B. (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory*, 32(2):359–401.

Wang, Q., Wu, D., and Zhu, K. (2018). Model checks for nonlinear cointegrating regression. *Journal of Econometrics*, 207(2):261–284.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92.

Supplementary Document for "Transformation models with cointegrated and deterministically trending regressors"

Yingqian Lin[a] and Yundong Tu[b*]

[a] School of Economics, Shanghai University of Finance and Economics

[b] Guanghua School of Management and

Center for Statistical Science, Peking University

October 14, 2022

This supplementary document provides the definitions on function classes, the technical details and additional simulation results for the paper "Transformation models with cointegrated and deterministically trending regressors". Appendix B presents the definitions of regular function and H-regular function. Appendix C contains some useful lemmas required in the proof of Theorem 1. Appendix D contains some additional simulation results.

Notations. $\mathbb{R}$ denotes the real line and $\mathbb{R}^+$ its positive part. For a vector $\boldsymbol{a} = (a_i)$ or a matrix $\boldsymbol{A} = (a_{ij})$, the norm $\|\cdot\|$ is defined as the maximum of the moduli such that $\|\boldsymbol{A}\| = \max_{i,j} |a_{i,j}|$ and $\|\boldsymbol{a}\| = \max_i |a_i|$. For a function f, which can be vector- or matrix-valued, $\|\cdot\|_K$ signifies the supremum norm over a subset K of its domain, that

*Corresponding author. Address: Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. E-mail: yundong.tu@gsm.pku.edu.cn.

is, $\|f\|_K = \sup_{x\in K}\|f(x)\|$. $\mathcal{F}_{LB}$ refers to the class of locally bounded functions on $\mathbb{R}$, the subclass $\mathcal{F}_{LB}^0$ of $\mathcal{F}_{LB}$ consists only of locally bounded functions that are exponentially bounded, i.e., functions f such that $f(x) = O(e^{c|x|})$ as $|x| \to \infty$ for some $c \in \mathbb{R}^+$. $\mathcal{F}_B$ signifies the class of bounded functions on $\mathbb{R}$ and its subclass $\mathcal{F}_B^0$ includes all functions that are bounded and vanish at infinity, i.e., functions f such that $f(x) \to 0$ as $|x| \to \infty$. Convergence in probability and convergence in distribution are denoted as $\xrightarrow{p}$ and $\Rightarrow$, respectively. Throughout this material, we denote generic constants by $C, C_1, C_2, \cdots$, which may be different at each appearance.

B Functional classes

Let $\mathcal{F}_{LB}^0$ be the class of locally bounded functions that are exponentially bounded, i.e., f fulfills condition $f(x) = O(e^{c|x|})$ as $|x| \to \infty$ for some $c \in \mathbb{R}^+$.

Definition B.1 *A function h is called* regular *on Θ if it satisfies the following conditions.*

(a) For all $\beta \in \Theta$, $h(\cdot, \beta)$ is continuous and twice differentiable;

(b) For all $x \in \mathbb{R}$, $h(x, \cdot)$ and $h^x(x, \cdot)$ are equicontinuous in a neighborhood of x, where $h^x(x, \cdot) = \partial h(x, \cdot)/\partial x$.

The continuity requirement on regular function h in Definition B.1 (a) is somewhat stronger than that defined in Definition 3.2 of Park and Phillips (2001). However, this condition is satisfied by most functions used in practical nonlinear time series analyses, including polynomial functions, logarithmic functions, etc. Naturally, a vector- or matrix-valued function is called regular when each of its components is regular.

Definition B.2 *A function $\boldsymbol{g}$ is called H-*regular *on Θ if it satisfies the following conditions.*

(a) $\boldsymbol{g}(\lambda x, \boldsymbol{\beta}) = \boldsymbol{\kappa}(\lambda, \boldsymbol{\beta})\boldsymbol{h}(x, \boldsymbol{\beta}) + \boldsymbol{R}(x, \lambda, \boldsymbol{\beta})$ *with* $\boldsymbol{\kappa}$ *nonsingular,* $\boldsymbol{h}$ *being regular on* Θ*, and* $\boldsymbol{R}(x, \lambda, \boldsymbol{\beta})$ *is differentiable with respect to* x*.* $\boldsymbol{R}(x, \lambda, \boldsymbol{\beta})$ *and its first-order partial derivative* $\boldsymbol{R}_1(x, \lambda, \boldsymbol{\beta}) \equiv \partial \boldsymbol{R}(x, \lambda, \boldsymbol{\beta})/\partial x$ *satisfy either*

(b.i) $|\boldsymbol{R}(x, \lambda, \boldsymbol{\beta})| \leq \boldsymbol{a}(\lambda, \boldsymbol{\beta})\boldsymbol{P}(x, \boldsymbol{\beta})$ *with* $\limsup_{\lambda\to\infty} \sup_{\boldsymbol{\beta}\in\Theta} |\boldsymbol{\kappa}^{-1}(\lambda, \boldsymbol{\beta})\boldsymbol{a}(\lambda, \boldsymbol{\beta})| = 0$*,* $|\boldsymbol{R}_1(x, \lambda, \boldsymbol{\beta})| \leq \boldsymbol{a}_1(\lambda, \boldsymbol{\beta})\boldsymbol{P}_1(x, \boldsymbol{\beta})$ *with* $\limsup_{\lambda\to\infty} \sup_{\boldsymbol{\beta}\in\Theta} |\boldsymbol{\kappa}^{-1}(\lambda, \boldsymbol{\beta})\boldsymbol{a}_1(\lambda, \boldsymbol{\beta})| = 0$*, or*

(b.ii) $|\boldsymbol{R}(x, \lambda, \boldsymbol{\beta})| \leq \boldsymbol{b}(\lambda, \boldsymbol{\beta})\boldsymbol{P}(x, \boldsymbol{\beta})\boldsymbol{Q}(\lambda x, \boldsymbol{\beta})$ *with* $\limsup_{\lambda\to\infty} \sup_{\boldsymbol{\beta}\in\Theta} |\boldsymbol{\kappa}^{-1}(\lambda, \boldsymbol{\beta})\boldsymbol{b}(\lambda, \boldsymbol{\beta})| < \infty$*,* $|\boldsymbol{R}_1(x, \lambda, \boldsymbol{\beta})| \leq \boldsymbol{b}_1(\lambda, \boldsymbol{\beta})\boldsymbol{P}_1(x, \boldsymbol{\beta})\boldsymbol{Q}_1(\lambda x, \boldsymbol{\beta})$ *with* $\limsup_{\lambda\to\infty} \sup_{\boldsymbol{\beta}\in\Theta} |\boldsymbol{\kappa}^{-1}(\lambda, \boldsymbol{\beta})\boldsymbol{b}_1(\lambda, \boldsymbol{\beta})| < \infty$*, where*

(c) $\sup_{\boldsymbol{\beta}\in\Theta} \boldsymbol{P}(\cdot, \boldsymbol{\beta}), \sup_{\boldsymbol{\beta}\in\Theta} \boldsymbol{P}_1(\cdot, \boldsymbol{\beta}) \in \mathcal{F}^0_{LB}$ *and* $\sup_{\boldsymbol{\beta}\in\Theta} \boldsymbol{Q}(\cdot, \boldsymbol{\beta}), \sup_{\boldsymbol{\beta}\in\Theta} \boldsymbol{Q}_1(\cdot, \boldsymbol{\beta}) \in \mathcal{F}^0_{B}$*.*

We call $\boldsymbol{\kappa}$ the asymptotic order, $\boldsymbol{h}$ the limit homogeneous function and $\boldsymbol{R}$ the remainder function of $\boldsymbol{g}$. Roughly speaking, the class of H-regular functions consists of functions that are asymptotically equivalent to their (uniquely defined) limit homogeneous functions. Condition (b) and (c) allow us to establish this asymptotic equivalence. The regularity requirement for the limit homogeneous function $\boldsymbol{h}$ in the condition (a) is necessary to ensure that $\boldsymbol{h}$ has well defined asymptotics. The regularity conditions in Definition B.1 and Definition B.2 are stronger than the corresponding conditions introduced in Park and Phillips (2001, Definition 3.2, Definition 3.5) and Uematsu (2019, Definition 7.1, Definition 7.2). In particular, smooth conditions on the regular function $\boldsymbol{h}$ and the remainder function $\boldsymbol{R}$ are imposed for technical convenience.

C Useful Lemmas

To develop the limit theory, we first rotate the integrated regressor $\boldsymbol{x}_t$ and the associated parameter $\boldsymbol{\theta}^0_1$ using an $(\ell_1 \times \ell_1)$-orthogonal matrix $\boldsymbol{M} = (\boldsymbol{m}_1, \boldsymbol{M}_2)$ with $\boldsymbol{m}_1 = \boldsymbol{\theta}^0_1/(\boldsymbol{\theta}^{0\prime}_1\boldsymbol{\theta}^0_1)^{1/2}$. The components $\boldsymbol{m}_1$ and $\boldsymbol{M}_2$ are of ranks 1 and $(\ell_1 - 1)$, respectively.

More explicitly, we have

$$\boldsymbol{M}'\boldsymbol{x}_t = \begin{pmatrix} \boldsymbol{m}_1'\boldsymbol{x}_t \\ \boldsymbol{M}_2'\boldsymbol{x}_t \end{pmatrix} \equiv \begin{pmatrix} \widetilde{x}_{1t} \\ \widetilde{\boldsymbol{x}}_{2t} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{M}'\boldsymbol{\theta}_1^0 = \begin{pmatrix} \boldsymbol{m}_1'\boldsymbol{\theta}_1^0 \\ \boldsymbol{M}_2'\boldsymbol{\theta}_1^0 \end{pmatrix} \equiv \begin{pmatrix} \alpha_{10} \\ \boldsymbol{\alpha}_{20} \end{pmatrix}, \tag{C.1}$$

where $\widetilde{x}_{1t}$, $\alpha_{10} = (\boldsymbol{\theta}_1^{0\prime}\boldsymbol{\theta}_1^0)^{1/2}$ are scalars, and $\widetilde{\boldsymbol{x}}_{2t}$, $\boldsymbol{\alpha}_{20} = 0$ are $(\ell_1 - 1)$-dimensional vectors. We accordingly define the limit Brownian Motions of x_{1t} and $\boldsymbol{x}_{2t}$ as

$$V_1 = \boldsymbol{m}_1'\boldsymbol{V} \qquad \text{and} \qquad \boldsymbol{V}_2 = \boldsymbol{M}_2'\boldsymbol{V},$$

which are of dimensions 1 and $(\ell_1 - 1)$, respectively.

For notational simplicity, define $\Lambda_t(\boldsymbol{\beta}) = \Lambda(y_t, \boldsymbol{\beta})$, $\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}) = \partial\Lambda_t^2(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$, $\boldsymbol{\kappa}_{n\boldsymbol{\xi},\boldsymbol{\beta}} = \boldsymbol{\kappa}_{\boldsymbol{\xi}}(\sqrt{n}, \boldsymbol{\beta})$, $\boldsymbol{\kappa}_{n\boldsymbol{\xi},0} = \boldsymbol{\kappa}_{\boldsymbol{\xi}}(\sqrt{n}, \boldsymbol{\beta}_0)$, $\boldsymbol{h}_{\boldsymbol{\xi},\boldsymbol{\beta}}(x) = \boldsymbol{h}_{\boldsymbol{\xi}}(x, \boldsymbol{\beta})$, $\boldsymbol{h}_{\boldsymbol{\xi},0}(x) = \boldsymbol{h}_{\boldsymbol{\xi}}(x, \boldsymbol{\beta}_0)$, $\boldsymbol{h}^x_{\boldsymbol{\xi},\boldsymbol{\beta}}(x) = \partial\boldsymbol{h}_{\boldsymbol{\xi}}(x, \boldsymbol{\beta})/\partial x$, $\boldsymbol{h}^x_{\boldsymbol{\xi},0}(x) = \boldsymbol{h}^x_{\boldsymbol{\xi},\boldsymbol{\beta}_0}(x)$, $\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}}(x, \lambda) = \boldsymbol{R}_{\boldsymbol{\xi}}(x, \lambda, \boldsymbol{\beta})$, $\boldsymbol{P}_{\boldsymbol{\xi},0}(x) = \boldsymbol{P}_{\boldsymbol{\xi}}(x, \boldsymbol{\beta}_0)$, $\boldsymbol{Q}_{\boldsymbol{\xi},0}(x) = \boldsymbol{Q}_{\boldsymbol{\xi}}(x, \boldsymbol{\beta}_0)$.

Lemma C.1 *Let Assumptions 1 hold. Then for any regular function f, we have*

(1) $\frac{1}{n}\sum_{t=1}^n f\left(\frac{\widetilde{x}_{1t}}{\sqrt{n}}, \boldsymbol{\theta}\right) \Rightarrow \int_0^1 f(V_1(r), \boldsymbol{\theta})dr$,

(2) $\frac{1}{n^{3/2}}\sum_{t=1}^n f\left(\frac{\widetilde{x}_{1t}}{\sqrt{n}}, \boldsymbol{\theta}\right)\widetilde{\boldsymbol{x}}_{2t} \Rightarrow \int_0^1 f(V_1(r), \boldsymbol{\theta})\boldsymbol{V}_2(r)dr$,

(3) $\frac{1}{n^2}\sum_{t=1}^n \widetilde{\boldsymbol{x}}_{2t}\widetilde{\boldsymbol{x}}_{2t}' \Rightarrow \int_0^1 \boldsymbol{V}_2(r)\boldsymbol{V}_2'(r)dr$,

(4) $\frac{1}{\sqrt{n}}\sum_{t=1}^n f\left(\frac{\widetilde{x}_{1t}}{\sqrt{n}}, \boldsymbol{\theta}\right)u_t \Rightarrow \int_0^1 f(V_1(r), \boldsymbol{\theta})dU(r)$,

(5) $\frac{1}{n}\sum_{t=1}^n \widetilde{\boldsymbol{x}}_{2t}u_t \Rightarrow \int_0^1 \boldsymbol{V}_2(r)dU(r)$,

(6) $\frac{1}{n}\sum_{t=1}^n f(\frac{\widetilde{x}_{1t}}{\sqrt{n}}, \boldsymbol{\theta})\boldsymbol{z}_t' \xrightarrow{p} 0$,

(7) $\frac{\boldsymbol{\kappa}_{nd}^{-1}}{n}\sum_{t=1}^n \boldsymbol{d}_t f(\frac{\widetilde{x}_{1t}}{\sqrt{n}}, \boldsymbol{\theta}) \Rightarrow \int_0^1 \boldsymbol{d}(r) f(V_1(r), \boldsymbol{\theta})dr$.

The weak convergence in (1)-(5) holds jointly.

Proof of Lemma C.1. Since $f(x, \boldsymbol{\theta})$ is continuous function, the result follows from Assumption 1 and the continuous mapping theorem. See, Kurtz and Protter (1991), Chang et al. (2001), Chan and Wang (2015), Hu et al. (2021) and Park and Phillips (2000) for instance.

Lemma C.2 *Define*

$$A_{n1}(\boldsymbol{\beta}) = \sum_{t=1}^{n} \Lambda_t^2(\boldsymbol{\beta}), \qquad A_{n5}(\boldsymbol{\beta}) = \sum_{t=1}^{n} \dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}),$$

$$A_{n2}(\boldsymbol{\beta}) = \sum_{t=1}^{n} [\Lambda_t(\boldsymbol{\beta}) - \boldsymbol{w}_t'\boldsymbol{\theta}_0]^2, \qquad A_{n6}(\boldsymbol{\beta}) = \sum_{t=1}^{n} [\Lambda_t(\boldsymbol{\beta}) - \boldsymbol{w}_t'\boldsymbol{\theta}_0]\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}),$$

$$A_{n3}(\boldsymbol{\beta}) = \sum_{t=1}^{n} \Lambda_t(\boldsymbol{\beta})\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}), \qquad A_{n7}(\boldsymbol{\beta}) = \sum_{t=1}^{n} \Lambda_t(\boldsymbol{\beta})\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}),$$

$$A_{n4}(\boldsymbol{\beta}) = \sum_{t=1}^{n} [\Lambda_t(\boldsymbol{\beta}) - \boldsymbol{w}_t'\boldsymbol{\theta}_0]\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}), \qquad A_{n8}(\boldsymbol{\beta}) = \sum_{t=1}^{n} \boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}).$$

Let Assumptions 1-3 hold.

(a) If $\kappa_{nd1}/\sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$, we have

(1) $n^{-2}A_{n1}(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \tau_3^2(r)dr$,

(2) $n^{-1}A_{n2}(\boldsymbol{\beta}_0) \xrightarrow{p} \sigma_u^2$,

(3) $(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \tau_3(r)\boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))dr$,

(4) $(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n4}(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))dU(r) + \sigma_u^2 \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}^x(\tau_3(r))dr$,

(5) $(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n5}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-\top} \Rightarrow \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_3(r))dr$,

(6) $(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n6}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-\top} = o_p(1)$,

(7) $(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n7}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top} = o_p(1)$,

(8) $n^{-3/2}A_{n8}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top} \Rightarrow \int_0^1 \boldsymbol{V}(r)\boldsymbol{h}_{\boldsymbol{\xi},0}^\top(\tau_3(r))dr$*; and*

(b) if $\sqrt{n}/\kappa_{nd1} \to 0$,

(1) $(n\kappa_{nd1}^2)^{-1}A_{n1}(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \tau_2^2(r)dr$,

(2) $n^{-1}A_{n2}(\boldsymbol{\beta}_0) \xrightarrow{p} \sigma_u^2$,

(3) $(n\kappa_{nd1}\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \tau_2(r)\boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_2(r))dr$,

(4) $(\sqrt{n}\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}A_{n4}(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_2(r))dU(r)$,

(5) $(\sqrt{n}\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}A_{n5}(\boldsymbol{\beta}_0)(\sqrt{n}\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-\top} \Rightarrow \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_2(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^{\top}(\tau_2(r))dr$,

(6) $(\sqrt{n}\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}A_{n6}(\boldsymbol{\beta}_0)(\sqrt{n}\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-\top} = o_p(1)$,

(7) $(n\kappa_{nd1}\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}A_{n7}(\boldsymbol{\beta}_0)\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-\top} = o_p(1)$,

(8) $n^{-3/2}A_{n8}(\boldsymbol{\beta}_0)\widetilde{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-\top} \Rightarrow \int_0^1 \boldsymbol{V}(r)\boldsymbol{h}_{\boldsymbol{\xi},0}^{\top}(\tau(r))dr$,

where $\tau_1(r) = \boldsymbol{V}(r)'\boldsymbol{\theta}_1^0$, $\tau_2(r) = d_1(r)\theta_{31}^0$, $\tau_3(r) = \tau_1(r) + \pi\tau_2(r)$, $\boldsymbol{\kappa}_{n\xi,0} = \boldsymbol{\kappa}_{\xi}(\sqrt{n}, \boldsymbol{\beta}_0)$, $\widetilde{\boldsymbol{\kappa}}_{n\xi,0} = \boldsymbol{\kappa}_{\xi}(\kappa_{nd1}, \boldsymbol{\beta}_0)$, $\boldsymbol{h}_{\boldsymbol{\xi},0}^{x} = \partial\boldsymbol{h}_{\boldsymbol{\xi},0}/\partial x$. *Here,* $d_1(r)$ *and* θ_{31}^0 *are the first elements of* $\boldsymbol{d}(r)$ *and* $\boldsymbol{\theta}_3^0$*, respectively. The weak convergence in (1)-(5) and (8) holds jointly.*

Proof of Lemma C.2 The arguments used to prove the two cases are similar, we present the proofs for Case (a): $\kappa_{nd1}/\sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$ below, with the detailed proofs for Case (b): $\sqrt{n}/\kappa_{nd1} \to 0$ omitted.

(a1) First write that

$$\frac{1}{n^2}A_{n1}(\boldsymbol{\beta}_0) = \frac{1}{n^2}\sum_{t=1}^{n}\Lambda_t^2(\boldsymbol{\beta}_0) = \frac{1}{n^2}\sum_{t=1}^{n}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0 + d_{1t}\theta_{31}^0)^2(1 + o_p(1)) \Rightarrow \int_0^1 \tau_3^2(r)dr,$$

by Lemma C.1.

(a2) The result follows from the law of large numbers.

(a3) First write that

$$\begin{aligned}\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}A_{n3}(\boldsymbol{\beta}_0) &= \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}\Lambda_t(\boldsymbol{\beta}_0)\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) \\ &= \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) + \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{z}_t'\boldsymbol{\theta}_2^0\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) \\ &+ \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{d}_t'\boldsymbol{\theta}_3^0\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) + \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}u_t\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0).\end{aligned}$$

By Assumption 3, we have

$$\begin{aligned}
\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\dot{\Lambda}_t(\boldsymbol{\beta}_0) &= \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\boldsymbol{\xi}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t,\boldsymbol{\beta}_0)\\
&= \frac{1}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)\\
&\quad+\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\\
&\equiv \Pi_{11}^a+\Pi_{12}^a,
\end{aligned}$$

where the definitions of Π_{11}^a-Π_{12}^a should be obvious. First consider Π_{11}^a, by the mean value theorem, we have

$$\Pi_{11}^a=\frac{1}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)+\frac{1}{n^2}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\boldsymbol{h}_{\boldsymbol{\xi},0}^x(\bar{x})(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t),$$

where $\bar{x}$ is between $\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}$ and $\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}$. By Lemma C.1 and the continuous mapping theorem, we have

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)\Rightarrow\int_0^1\tau_1(r)\boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dr.$$

Using arguments similar to Lemma A4 in Park and Phillips (2001), first let $s_{\max}=\max_{0\le r\le 1}\tau_3(r)$ and $s_{\min}=\min_{0\le r\le 1}\tau_3(r)$ and $s_m=\max(s_{\max},-s_{\min})+1$. Since $s_m<\infty$, a.s., we have $P\{s_m>C\}\to 0$ as $C\to\infty$. There we may take $C>0$ large so that $P\{s_m>C\}$ is arbitrarily small. In addition, since $(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)/\sqrt{n}\xrightarrow{p}0$, there exist a large constant $C_1>0$ such that $P\{|(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)/\sqrt{n}|>C_1\}\to 0$. Fix $C,C_1>0$ large, and define $K=[-(C+C_1),C+C_1]$. Therefore, we have

$$\begin{aligned}
&\Big\|\frac{1}{n^2}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{\theta}_1^0\boldsymbol{h}_{\boldsymbol{\xi},0}^x(\bar{x})(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)\Big\|\\
&\le\|\boldsymbol{h}_{\boldsymbol{\xi},0}^x\|_K\frac{1}{n^2}\sum_{t=1}^{n}|\boldsymbol{x}_t'\boldsymbol{\theta}_1^0(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)|\\
&\le\|\boldsymbol{h}_{\boldsymbol{\xi},0}^x\|_K\frac{1}{\sqrt{n}}\Big(\frac{1}{n^2}\sum_{t=1}^{n}[\boldsymbol{x}_t'\boldsymbol{\theta}_1^0]^2\Big)^{1/2}\Big(\frac{1}{n}\sum_{t=1}^{n}[\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t]^2\Big)^{1/2}=o_p(1),
\end{aligned}$$

and

$$\Pi_{11}^{a} \Rightarrow \int_0^1 \tau_1(r)\boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dr. \tag{C.2}$$

Next for Π_{12}^{a}, by Assumption 3, $\boldsymbol{P}_{\boldsymbol{\xi},0} \in \mathcal{F}_{LB}^{0}$, $\boldsymbol{Q}_{\boldsymbol{\xi},0} \in \mathcal{F}_{B}^{0}$, and we have

$$\begin{aligned}
\|\Pi_{12}^{a}\| &= \Big\| \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \boldsymbol{x}_t'\boldsymbol{\theta}_1^0 \boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0 + \boldsymbol{z}_t'\boldsymbol{\theta}_2^0 + \boldsymbol{d}_t'\boldsymbol{\theta}_3^0 + u_t}{\sqrt{n}}, \sqrt{n}\Big)\Big\| \\
(i) &\leq \|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K \cdot \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{a}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\| \cdot \frac{1}{n^{3/2}}\sum_{t=1}^{n}|\boldsymbol{x}_t'\boldsymbol{\theta}_1^0| = o_p(1), \\
(ii) &\leq \Big\|\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0}{\sqrt{n}}\Big\|_K \cdot \|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K \cdot \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{b}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\| \cdot \frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{Q}_{\boldsymbol{\xi},0}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0 + \boldsymbol{z}_t'\boldsymbol{\theta}_2^0 + \boldsymbol{d}_t'\boldsymbol{\theta}_3^0 + u_t)\| = o_p(1).
\end{aligned}$$

Above, we use (i) to signify that the function satisfies (b.i) in Definition B.2, and use (ii) for (b.ii) analogously. Similar notations are used throughout the proof. Thus,

$$\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \boldsymbol{x}_t'\boldsymbol{\theta}_1^0 \dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \tau_1(r)\boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dr.$$

Similarly, we have

$$\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \boldsymbol{d}_t'\boldsymbol{\theta}_3^0 \dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \pi\tau_2(r)\boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dr.$$

And by Assumption 1 and 3,

$$\begin{aligned}
\Big\| \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \boldsymbol{z}_t'\boldsymbol{\theta}_2^0 \dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\Big\| &\leq \frac{1}{n^{3/2}}\sum_{t=1}^{n}|\boldsymbol{z}_t'\boldsymbol{\theta}_2^0| \cdot \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\| \\
&\leq \frac{1}{\sqrt{n}}\Big(\frac{1}{n}\sum_{t=1}^{n}|\boldsymbol{z}_t'\boldsymbol{\theta}_2^0|^2\Big)^{1/2}\Big(\frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\|^2\Big)^{1/2} = o_p(1),
\end{aligned}$$

by Lemma C.2 (a4), we have $(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\sum_{t=1}^{n} u_t\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) = o_p(1)$. Thus,

$$\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} A_{n3}(\boldsymbol{\beta}_0) = \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0 + \boldsymbol{d}_t'\boldsymbol{\theta}_3^0)\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) + o_p(1) \Rightarrow \int_0^1 \tau_3(r)\boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dr.$$

(a4) By Assumption 3, write

$$
\begin{aligned}
\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}u_t\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) &= \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}u_t\boldsymbol{\xi}\big(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t,\boldsymbol{\beta}_0\big)\\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)\\
&+\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}u_t\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\\
&\equiv \Pi_{21}^b+\Pi_{22}^b,
\end{aligned}
$$

where the definitions of Π_{21}^b and Π_{22}^b should be obvious. By the Taylor expansion,

$$
\begin{aligned}
\Pi_{21}^b = &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)+\frac{1}{n}\sum_{t=1}^{n}u_t^2\boldsymbol{h}_{\boldsymbol{\xi},0}^x\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)\\
&+\frac{1}{n}\sum_{t=1}^{n}u_t\boldsymbol{z}_t'\boldsymbol{\theta}_2^0\boldsymbol{h}_{\boldsymbol{\xi},0}^x\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)+\frac{1}{n^{3/2}}\sum_{t=1}^{n}u_t(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)^2\boldsymbol{h}_{\boldsymbol{\xi},0}^{xx}(\bar{x}),
\end{aligned}
$$

where $\bar{x}$ is between $\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}$ and $\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}$. Note that by Lemma C.1 and the continuous mapping theory, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)\Rightarrow\int_0^1\boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dU(r),\\
&\frac{1}{n}\sum_{t=1}^{n}u_t^2\boldsymbol{h}_{\boldsymbol{\xi},\boldsymbol{\beta}}^x\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)=\sigma_u^2\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{h}_{\boldsymbol{\xi},\boldsymbol{\beta}}^x\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)+\frac{1}{n}\sum_{t=1}^{n}(u_t^2-\sigma_u^2)\boldsymbol{h}_{\boldsymbol{\xi},\boldsymbol{\beta}}^x\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\Rightarrow\sigma_u^2\int_0^1\boldsymbol{h}_{\boldsymbol{\xi},\boldsymbol{\beta}}^x\big(\tau_3(r)\big)dr.
\end{aligned}
$$

It is easy to show $\frac{1}{n}\sum_{t=1}^n u_t\boldsymbol{z}_t'\boldsymbol{\theta}_2^0\boldsymbol{h}_{\boldsymbol{\xi},0}^x\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)=o_p(1)$, and $\frac{1}{n^{3/2}}\sum_{t=1}^n u_t(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)^2\boldsymbol{h}_{\boldsymbol{\xi},0}^{xx}(\bar{x})=o_p(1)$. Thus, $\Pi_{21}^b\Rightarrow\int_0^1\boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dU(r)+\sigma_u^2\int_0^1\boldsymbol{h}_{\boldsymbol{\xi},\boldsymbol{\beta}}^x\big(\tau_3(r)\big)dr$. For Π_{22}^b, by the mean value theorem, we have

$$
\begin{aligned}
\Pi_{22}^b &= \frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}u_t\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}},\sqrt{n}\Big)+\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n}\sum_{t=1}^{n}u_t(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}^x(\bar{x},\sqrt{n})\\
&\equiv \Pi_{221}^b+\Pi_{222}^b,
\end{aligned}
$$

where $\bar{x}$ is between $\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}$ and $\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}$ and the definitions of Π_{221}^b, Π_{222}^b should be obvious. Using arguments similar to Lemma A4 in Park and Phillips (2001), write that

$$\|E[\Pi_{221}^b{\Pi_{221}^b}']\| = \Big\|\sigma_u^2(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}E\Big[\sum_{t=1}^n \boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}},\sqrt{n}\Big)\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}^\top\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}},\sqrt{n}\Big)\Big](\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-T}\Big\|$$

$$(i) \le \sigma_u^2\cdot\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{a}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|^2\cdot E\Big[\frac{1}{n}\sum_{t=1}^n\Big\|\boldsymbol{P}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)\Big\|^2\Big]$$

$$\le \sigma_u^2\cdot\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{a}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|^2\times E\|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K^2 = o_p(1),$$

$$(ii) \le \sigma_u^2\cdot\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{b}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|^2\times E\Big[\frac{1}{n}\sum_{t=1}^n\Big\|\boldsymbol{P}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0}{\sqrt{n}}\Big)\boldsymbol{Q}_{\boldsymbol{\xi},0}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0)\Big\|^2\Big]$$

$$\le \sigma_u^2\cdot\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{b}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|^2\times E\|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K^2\frac{1}{n}\sum_{t=1}^n\|\boldsymbol{Q}_{\boldsymbol{\xi},0}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0)\|^2 = o_p(1),$$

$$\|\Pi_{222}^b\| \quad (i) \le \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{a}_{\boldsymbol{\xi}}^x(\sqrt{n},\boldsymbol{\beta}_0)\|\cdot\|\boldsymbol{P}_{\boldsymbol{\xi},0}^x\|_K\cdot\frac{1}{n}\sum_{t=1}^n u_t(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t) = o_p(1),$$

$$(ii) \le \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{b}_{\boldsymbol{\xi}}^x(\sqrt{n},\boldsymbol{\beta}_0)\|\times\frac{1}{n}\sum_{t=1}^n u_t(\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+u_t)\boldsymbol{P}_{\boldsymbol{\xi},0}^x(\bar{x})\boldsymbol{Q}_{\boldsymbol{\xi},0}^x(\sqrt{n}\bar{x}) = o_p(1).$$

Thus, $\Pi_{22}^b = o_p(1)$, and $(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n4}(\boldsymbol{\beta}_0) \Rightarrow \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}\big(\tau_3(r)\big)dU(r)+\sigma_u^2\int_0^1\boldsymbol{h}_{\boldsymbol{\xi},\boldsymbol{\beta}}^x\big(\tau_3(r)\big)dr$.

(a5) Similarly,

$$\begin{aligned}
&\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}A_{n5}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-T}\\
&=\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}\sum_{t=1}^n\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-T}\\
&=\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}\sum_{t=1}^n\Big[\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)+\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\Big]\\
&\quad\times\Big[\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)+\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\Big]^\top\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}}{\sqrt{n}}\\
&=\frac{1}{n}\sum_{t=1}^n\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)\boldsymbol{h}_{\boldsymbol{\xi},0}^\top\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)\\
&+\frac{1}{n}\sum_{t=1}^n\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{R}_{\boldsymbol{\xi},\boldsymbol{\beta}_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\boldsymbol{h}_{\boldsymbol{\xi},0}^\top\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{h}_{\boldsymbol{\xi},0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}}\Big)\boldsymbol{R}_{\boldsymbol{\xi},\beta_0}^{\top}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\\
&+\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{R}_{\boldsymbol{\xi},\beta_0}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\boldsymbol{R}_{\boldsymbol{\xi},\beta_0}^{\top}\Big(\frac{\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t}{\sqrt{n}},\sqrt{n}\Big)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\\
&\equiv \Pi_{31}^c+\Pi_{32}^c+{\Pi_{32}^c}^{\top}+\Pi_{33}^c,
\end{aligned}
$$

where the definitions of Π_{31}^c-Π_{33}^c should be obvious. Using arguments similar to (a3), we have

$$
\Pi_{31}^c \Rightarrow \int_0^1 \boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^{\top}(\tau_3(r))dr.
$$

By Assumption 3, we have

$$
\begin{aligned}
\|\Pi_{32}^c\|(i) &\le \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{a}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|\cdot\|\boldsymbol{h}_{\boldsymbol{\xi},0}\|_K\|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K=o_p(1),\\
(ii) &\le \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{b}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|\cdot\|\boldsymbol{h}_{\boldsymbol{\xi},0}\|_K\|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K\cdot\frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{Q}_{\boldsymbol{\xi},0}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t)\|=o_p(1),\\
\|\Pi_{33}^c\|(i) &\le \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{a}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|^2\cdot\|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K^2=o_p(1),\\
(ii) &\le \|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\boldsymbol{b}_{\boldsymbol{\xi}}(\sqrt{n},\boldsymbol{\beta}_0)\|^2\cdot\|\boldsymbol{P}_{\boldsymbol{\xi},0}\|_K^2\cdot\frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{Q}_{\boldsymbol{\xi},0}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0+\boldsymbol{z}_t'\boldsymbol{\theta}_2^0+\boldsymbol{d}_t'\boldsymbol{\theta}_3^0+u_t)\|^2=o_p(1).
\end{aligned}
$$

Thus,

$$
\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{\sqrt{n}}A_{n5}(\boldsymbol{\beta}_0)(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-T}=\Pi_{31}^c+o_p(1)\Rightarrow\int_0^1\boldsymbol{h}_{\boldsymbol{\xi},0}(\tau_3(r))\boldsymbol{h}_{\boldsymbol{\xi},0}^{\top}(\tau_3(r))dr.
$$

(a6) It follows from (3.2) that

$$
\begin{aligned}
\Big\|\frac{1}{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n6}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\Big\| &= \Big\|\frac{1}{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\sum_{t=1}^{n}\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)u_t\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\Big\|\\
&\le \Big\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\Big(\sup_{|s|\le\bar{s}}\dot{\boldsymbol{\xi}}(\sqrt{n}s,\boldsymbol{\beta}_0)\Big)\Big\|\cdot\frac{1}{n}\sum_{t=1}^{n}|u_t|=o_p(1).
\end{aligned}
$$

(a7) It follows from (3.2) that

$$
\begin{aligned}
&\left\|\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} A_{n7}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\| \\
&= \left\|\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \Lambda_t(\boldsymbol{\beta}_0)\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\| \\
&\leq \left\|\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \boldsymbol{x}_t'\boldsymbol{\theta}_1^0\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\| + \left\|\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \boldsymbol{z}_t'\boldsymbol{\theta}_2^0\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\| \\
&+ \left\|\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} \boldsymbol{d}_t'\boldsymbol{\theta}_3^0\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\| + \left\|\frac{\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}}{n^{3/2}} \sum_{t=1}^{n} u_t\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\| \\
&\leq \left\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\Big(\sup_{|s|\leq\bar{s}}\dot{\boldsymbol{\xi}}(\sqrt{n}s,\boldsymbol{\beta}_0)\Big)\right\| \cdot \frac{1}{n^{3/2}}\sum_{t=1}^{n}[|\boldsymbol{x}_t'\boldsymbol{\theta}_1^0| + |\boldsymbol{z}_t'\boldsymbol{\theta}_2^0| + |\boldsymbol{d}_t'\boldsymbol{\theta}_3^0| + |u_t|] = o_p(1).
\end{aligned}
$$

(a8) It follows from Lemma C.1 and arguments similar to (a3).

The results in (b) could be obtained using arguments similar to (a). The details are omitted to avoid repetition.

Lemma C.3 *Define* $\varpi_n = \max\{\sqrt{n}, \kappa_{nd1}\}$, $\overline{\boldsymbol{\kappa}}_{n\xi,0} = \boldsymbol{\kappa}_{\xi}(\varpi_n, \boldsymbol{\beta}_0)$, $\boldsymbol{C}_n = n^{1/2-\rho}\overline{\boldsymbol{\kappa}}_{n\xi,0}$, *where* $0 < \rho < \varepsilon/6$ *with* ε *defined in Assumption 3 and* $N_n = \{\boldsymbol{\beta} : \|\boldsymbol{C}_n(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \leq 1\}$. *Let Assumptions 1-3 hold. As* $n \to \infty$, *we have*

(1) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}\varpi_n^{-2}[A_{n1}(\boldsymbol{\beta}) - A_{n1}(\boldsymbol{\beta}_0)]\| = o_p(1)$,

(2) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}[A_{n2}(\boldsymbol{\beta}) - A_{n2}(\boldsymbol{\beta}_0)]\| = o_p(1)$,

(3) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}\varpi_n^{-1}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-1}[A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0)]\| = o_p(1)$,

(4) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-1}[A_{n4}(\boldsymbol{\beta}) - A_{n4}(\boldsymbol{\beta}_0)]\| = o_p(1)$,

(5) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta}) - A_{n5}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1)$,

(6) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-1}[A_{n6}(\boldsymbol{\beta}) - A_{n6}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1)$,

(7) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-1}\varpi_n^{-1}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-1}[A_{n7}(\boldsymbol{\beta}) - A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1)$.

In addition, we have

(i) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{-1}\varpi_n^{-2}A_{n1}(\boldsymbol{\beta})\| = O_p(1)$,

(ii) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{-1}A_{n2}(\boldsymbol{\beta})\| = O_p(1)$,

(iii) $\sup_{\boldsymbol{\beta}\in N_n} \|(n\varpi_n\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta})\| = O_p(1)$,

(iv) $\sup_{\boldsymbol{\beta}\in N_n} \|(\sqrt{n}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}A_{n5}(\boldsymbol{\beta})(\sqrt{n}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0})^{-1}\| = O_p(1)$,

(v) $\sup_{\boldsymbol{\beta}\in N_n} \|n^{-1}\varpi_n^{-1}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-1}A_{n7}(\boldsymbol{\beta})\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},0}^{-\top}\| = o_p(1)$.

Proof of Lemma C.3. The arguments used to prove the two cases are similar, we present the proofs for Case (a): $\kappa_{nd1}/\sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$ below, with the detailed proofs for Case (b): $\sqrt{n}/\kappa_{nd1} \to 0$ omitted.

Let $\bar{s} = s_m + C_1 + 1$, where s_m, C_1 are defined in the proof of Lemma C.2. Then we have for large n,

$$\begin{aligned}
\sup_{\boldsymbol{\beta}\in N_n} |\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})| &= \sup_{\boldsymbol{\beta}\in N_n} |\boldsymbol{\xi}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0 + \boldsymbol{z}_t'\boldsymbol{\theta}_2^0 + \boldsymbol{d}_t'\boldsymbol{\theta}_3^0 + u_t, \boldsymbol{\beta})| \le \sup_{|s|\le\bar{s}} \sup_{\boldsymbol{\beta}\in N_n} |\boldsymbol{\xi}(\sqrt{n}s, \boldsymbol{\beta})|,\\
\sup_{\boldsymbol{\beta}\in N_n} |\ddot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})| &= \sup_{\boldsymbol{\beta}\in N_n} |\dot{\boldsymbol{\xi}}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0 + \boldsymbol{z}_t'\boldsymbol{\theta}_2^0 + \boldsymbol{d}_t'\boldsymbol{\theta}_3^0 + u_t, \boldsymbol{\beta})| \le \sup_{|s|\le\bar{s}} \sup_{\boldsymbol{\beta}\in N_n} |\dot{\boldsymbol{\xi}}(\sqrt{n}s, \boldsymbol{\beta})|,\\
\sup_{\boldsymbol{\beta}\in N_n} |\dddot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})| &= \sup_{\boldsymbol{\beta}\in N_n} |\ddot{\boldsymbol{\xi}}(\boldsymbol{x}_t'\boldsymbol{\theta}_1^0 + \boldsymbol{z}_t'\boldsymbol{\theta}_2^0 + \boldsymbol{d}_t'\boldsymbol{\theta}_3^0 + u_t, \boldsymbol{\beta})| \le \sup_{|s|\le\bar{s}} \sup_{\boldsymbol{\beta}\in N_n} |\ddot{\boldsymbol{\xi}}(\sqrt{n}s, \boldsymbol{\beta})|.
\end{aligned}$$

(1) First write that

$$A_{n1}(\boldsymbol{\beta}) - A_{n1}(\boldsymbol{\beta}_0) = \sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2 + 2\sum_{t=1}^{n}\Lambda_t(\boldsymbol{\beta}_0)[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)].$$

It follows from the mean value theorem and Assumption 3 that

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta}) - A_{n1}(\boldsymbol{\beta}_0)]\|\\
&\le \sup_{\boldsymbol{\beta}\in N_n} \left\|n^{2\rho-2}\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2\right\| + 2\sup_{\boldsymbol{\beta}\in N_n} \left\|n^{2\rho-2}\sum_{t=1}^{n}\Lambda_t(\boldsymbol{\beta}_0)[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]\right\|\\
&\le n^{-1-2\rho}\cdot n^{6\rho-1}\left\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Big(\sup_{s\le\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\boldsymbol{\xi}(\sqrt{n}s, \boldsymbol{\beta})|\Big)\right\|^2\\
&+ 2n^{-1/2}\cdot\sqrt{n}^{6\rho-1}\left\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Big(\sup_{|s|\le\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\boldsymbol{\xi}(\sqrt{n}s, \boldsymbol{\beta})|\Big)\right\|\cdot\frac{1}{n^{3/2}}\sum_{t=1}^{n}|\Lambda_t(\boldsymbol{\beta}_0)| = o_p(1).
\end{aligned}$$

(2) Similarly, we write

$$A_{n2}(\boldsymbol{\beta}) - A_{n2}(\boldsymbol{\beta}_0) = \sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2 + 2\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]u_t.$$

By the mean value theorem and Assumption 3, we have

$$\begin{aligned}
\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{2\rho-1}\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2\right\| &\leq n^{-2\rho}\cdot n^{6\rho-1}\left\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Big(\sup_{s\leq\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\boldsymbol{\xi}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\|^2 = o_p(1),\\
\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{2\rho-1}\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]u_t\right\| &\leq n^{3\rho-3/2}\sup_{\boldsymbol{\beta}\in N_n}\sum_{t=1}^{n}\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\bar{\boldsymbol{\beta}})\||u_t|\\
&\leq \sqrt{n}^{6\rho-1}\left\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Big(\sup_{s\leq\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\boldsymbol{\xi}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\|\cdot\frac{1}{n}\sum_{t=1}^{n}|u_t| = o_p(1).
\end{aligned}$$

(3) Write

$$A_{n3}(\boldsymbol{\beta}) - A_{n3}(\boldsymbol{\beta}_0) = \sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}) + \sum_{t=1}^{n}\Lambda_t(\boldsymbol{\beta}_0)[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}) - \dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)].$$

Note that we have

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})\right\|\\
&\leq n^{3\rho-2}\sum_{t=1}^{n}\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\|\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})\|\\
&+\frac{1}{2}n^{4\rho-5/2}\sum_{t=1}^{n}\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\ddot{\boldsymbol{\Lambda}}_t(\bar{\boldsymbol{\beta}})\|\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})\|\\
&\leq n^{-1/2}\cdot\sqrt{n}^{6\rho-1}\left\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Big(\sup_{s\leq\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\boldsymbol{\xi}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\|\cdot\frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\|\\
&+\frac{1}{2}n^{-1/2-2\rho}\cdot\sqrt{n}^{6\rho-1}\left\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\Big(\sup_{s\leq\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\dot{\boldsymbol{\xi}}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\|\\
&\quad\times\sqrt{n}^{6\rho-1}\left\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Big(\sup_{s\leq\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\boldsymbol{\xi}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\| = o_p(1),
\end{aligned}$$

by the Taylor expansion, Assumption 3, and

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n} \|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\sum_{t=1}^{n}\Lambda_t(\boldsymbol{\beta}_0)[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)]\| \\
&\le n^{3\rho-2}\sup_{\boldsymbol{\beta}\in N_n}\sum_{t=1}^{n}|\Lambda_t(\boldsymbol{\beta}_0)|\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\ddot{\boldsymbol{\Lambda}}_t(\bar{\boldsymbol{\beta}})\| \\
&\le \sqrt{n}^{6\rho-1}\left\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\left(\sup_{s\le\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\dot{\boldsymbol{\xi}}(\sqrt{n}s,\boldsymbol{\beta})|\right)\right\|\cdot\frac{1}{n^{3/2}}\sum_{t=1}^{n}|\Lambda_t(\boldsymbol{\beta}_0)| = o_p(1),
\end{aligned}
$$

by the mean value theorem and Assumption 3. Thus, $\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)]\| = o_p(1)$.

(4) First write

$$
\begin{aligned}
A_{n4}(\boldsymbol{\beta})-A_{n4}(\boldsymbol{\beta}_0) &= \sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta})-\Lambda_t(\boldsymbol{\beta}_0)][\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)]+\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta})-\Lambda_t(\boldsymbol{\beta}_0)]\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0) \\
&\quad+\sum_{t=1}^{n}[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)]u_t.
\end{aligned}
$$

And the result follows from arguments similar to (3).

(5) Following arguments similar to that of the proof of (1)-(3), the proof can be completed by writing

$$
\begin{aligned}
A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0) &= \sum_{t=1}^{n}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)]^\top+\sum_{t=1}^{n}[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)][\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)]^\top \\
&\quad+\sum_{t=1}^{n}[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)]\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_0),
\end{aligned}
$$

the details of which are omitted here.

(6) Note that

$$
A_{n6}(\boldsymbol{\beta})-A_{n6}(\boldsymbol{\beta}_0) = \sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta})-\Lambda_t(\boldsymbol{\beta}_0)]\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta})+\sum_{t=1}^{n}[\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta})-\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)]u_t.
$$

By the Taylor expansion and (3.4), we have

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta})-\Lambda_t(\boldsymbol{\beta}_0)]\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta})\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\|\\
&\le n^{3\rho-3/2}\sup_{\boldsymbol{\beta}\in N_n}\sum_{t=1}^{n}\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\|\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1})^{-1}\ddot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})\|\\
&+\frac{1}{2}n^{4\rho-2}\sup_{\boldsymbol{\beta}\in N_n}\sum_{t=1}^{n}\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1})^{-1}\ddot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta})\|\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1})^{-1}\ddot{\boldsymbol{\Lambda}}_t(\bar{\boldsymbol{\beta}})\|\\
&\le\sqrt{n}^{6\rho-1}\left\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\Big(\sup_{s\le\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\dot{\boldsymbol{\xi}}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\|\cdot\frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\|\\
&+\frac{1}{2}n^{-2\rho}\cdot n^{6\rho-1}\left\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\Big(\sup_{s\le\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\dot{\boldsymbol{\xi}}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\|^2=o_p(1).
\end{aligned}$$

And by the mean value theorem and (3.5), we have

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\sum_{t=1}^{n}[\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta})-\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)]u_t\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-T}\right\|\\
&\le n^{3\rho-3/2}\sup_{\boldsymbol{\beta}\in N_n}\sum_{t=1}^{n}\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\dddot{\boldsymbol{\Lambda}}_t(\bar{\boldsymbol{\beta}})\||u_t|\\
&\le\sqrt{n}^{6\rho-1}\left\|(\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\otimes\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}\Big(\sup_{s\le\bar{s}}\sup_{\boldsymbol{\beta}\in N_n}|\ddot{\boldsymbol{\xi}}(\sqrt{n}s,\boldsymbol{\beta})|\Big)\right\|\cdot\frac{1}{n}\sum_{t=1}^{n}|u_t|=o_p(1).
\end{aligned}$$

Thus, $\sup_{\boldsymbol{\beta}\in N_n}\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n6}(\boldsymbol{\beta})-A_{n6}(\boldsymbol{\beta}_0)]\|=o_p(1)$.

(7) First write,

$$\begin{aligned}
A_{n7}(\boldsymbol{\beta})-A_{n7}(\boldsymbol{\beta}_0)&=\sum_{t=1}^{n}\Lambda_t(\boldsymbol{\beta}_0)[\ddot{\lambda}_t(\boldsymbol{\beta})-\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)]+\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta})-\Lambda_t(\boldsymbol{\beta}_0)][\ddot{\lambda}_t(\boldsymbol{\beta})-\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0)]\\
&+\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta})-\Lambda_t(\boldsymbol{\beta}_0)]\ddot{\boldsymbol{\lambda}}_t(\boldsymbol{\beta}_0).
\end{aligned}$$

The result follows from an argument similar to (3) and (6).

The results in (i)-(v) follow from Lemma C.3 (1)-(7) and Lemma C.2 directly.

Lemma C.4 *Let Assumptions 1-3 hold. For each $\boldsymbol{\beta}$ satisfying $\|\sqrt{n}\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},\boldsymbol{\beta}_0}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|\le C$, where $\overline{\boldsymbol{\kappa}}_{n\boldsymbol{\xi},\boldsymbol{\beta}_0}$ is defined in Lemma C.3, we have*

$$L_n(\boldsymbol{\beta})=\frac{\sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta})-\boldsymbol{w}_t'\boldsymbol{\theta}_0]^2}{\sum_{t=1}^{n}\Lambda_t^2(\boldsymbol{\beta})}(1+o_p(1)). \tag{C.3}$$

Proof of Lemma C.4 The arguments used to prove the two cases are similar, we present the proofs for Case (a): $\kappa_{nd1}/\sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$ below, with the detailed proofs for Case (b): $\sqrt{n}/\kappa_{nd1} \to 0$ omitted.

Write that

$$
\begin{aligned}
L_n(\boldsymbol{\beta}) &= \frac{\sum_{t=1}^n [\Lambda_t(\boldsymbol{\beta}) - \boldsymbol{w}_t'\boldsymbol{\theta}_0 + \boldsymbol{w}_t'\boldsymbol{\theta}_0 - \boldsymbol{w}_t'\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})]^2}{\sum_{t=1}^n \Lambda_t^2(\boldsymbol{\beta})} \\
&= \frac{\sum_{t=1}^n [\Lambda_t(\boldsymbol{\beta}) - \boldsymbol{w}_t'\boldsymbol{\theta}_0]^2 + \sum_{t=1}^n [\boldsymbol{w}_t'\boldsymbol{\theta}_0 - \boldsymbol{w}_t'\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})]^2 + 2\sum_{t=1}^n [\Lambda_t(\boldsymbol{\beta}) - \boldsymbol{w}_t'\boldsymbol{\theta}_0][\boldsymbol{w}_t'\boldsymbol{\theta}_0 - \boldsymbol{w}_t'\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})]}{\sum_{t=1}^n \Lambda_t^2(\boldsymbol{\beta})} \\
&\equiv \frac{L_{n1} + L_{n2} + 2L_{n3}}{L_{n0}},
\end{aligned}
$$

where the definitions of L_{n0}-L_{n3} should be obvious. Consider L_{n1} first, we have

$$
\begin{aligned}
L_{n1} &= \sum_{t=1}^n [\Lambda_t(\boldsymbol{\beta}) - \boldsymbol{w}_t'\boldsymbol{\theta}_0]^2 = \sum_{t=1}^n [u_t + \Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2 \\
&= \sum_{t=1}^n u_t^2 + 2\sum_{t=1}^n u_t[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)] + \sum_{t=1}^n [\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2.
\end{aligned}
$$

By the law of large number, we have $\sum_{t=1}^n u_t^2 = O_p(n)$. Using arguments similar to Lemma C.3, we have

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}\in N_n} \left\| n^{-1}\sum_{t=1}^n [\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]u_t \right\| \le n^{-1/2} \left\| \boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1} \Big(\sup_{s\le\bar{s}} \sup_{\boldsymbol{\beta}\in N_n} |\boldsymbol{\xi}(\sqrt{n}s, \boldsymbol{\beta})| \Big) \right\| \cdot \frac{1}{n}\sum_{t=1}^n |u_t| = o_p(1), \\
&\sup_{\boldsymbol{\beta}\in N_n} \left\| n^{-1}\sum_{t=1}^n [\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2 \right\| \le n^{-1} \left\| \boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1} \Big(\sup_{s\le\bar{s}} \sup_{\boldsymbol{\beta}\in N_n} |\boldsymbol{\xi}(\sqrt{n}s, \boldsymbol{\beta})| \Big) \right\|^2 = o_p(1),
\end{aligned}
$$

by the mean value theorem and Assumption 3. Therefore, $L_{n1} = O_p(n)$. To show (C.3), we just need to show L_{n2}-$L_{n3} = o_p(n)$.

For L_{n2}, we have

$$
\begin{aligned}
L_{n2} &= \sum_{t=1}^n [\boldsymbol{w}_t'\boldsymbol{\theta}_0 - \boldsymbol{w}_t'\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})]^2 = \sum_{t=1}^n \left[\boldsymbol{w}_t'(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}_0)) + \boldsymbol{w}_t'(\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}_0) - \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}))\right]^2 \\
&\le 2\sum_{t=1}^n \left\{ \left[\boldsymbol{w}_t'(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}_0))\right]^2 + \left[\boldsymbol{w}_t'(\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}_0) - \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}))\right]^2 \right\} \\
&\equiv 2(L_{n21} + L_{n22}), \qquad \text{(C.4)}
\end{aligned}
$$

where the definitions of L_{n21}-L_{n22} should be obvious. For L_{n21}, by Lemma C.1,

$$\begin{aligned} L_{n21} &= \boldsymbol{U}'\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{U} \\ &= \boldsymbol{U}'\boldsymbol{W}(\boldsymbol{D}\boldsymbol{D}^{-1}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{D}^{-1}\boldsymbol{D})^{-1}\boldsymbol{W}'\boldsymbol{U} \\ &\leq \lambda_{\min}^{-1}(\boldsymbol{D}^{-1}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{D}^{-1}) \cdot \boldsymbol{U}'\boldsymbol{W}\boldsymbol{D}^{-2}\boldsymbol{W}'\boldsymbol{U} = o_p(n), \end{aligned}$$

where $\boldsymbol{D} = diag\{n\boldsymbol{I}_{\ell_1}, \sqrt{n}\boldsymbol{I}_{\ell_2}, \sqrt{n}\boldsymbol{\kappa}_{nd}\}$, $\lambda_{\min}(\boldsymbol{A})$ is the smallest eigenvalue of matrix $\boldsymbol{A}$ and $\boldsymbol{U} = (u_1, \cdots, u_n)'$. For L_{n22}, we have

$$\begin{aligned} L_{n22} &= (\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}_0) - \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}))^\top \boldsymbol{W}^\top \boldsymbol{W}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}_0) - \widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})) \\ &= (\Lambda(\boldsymbol{\beta}) - \Lambda(\boldsymbol{\beta}_0))^\top \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top(\Lambda(\boldsymbol{\beta}) - \Lambda(\boldsymbol{\beta}_0)) \\ &\leq \lambda_{\max}\big(\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\big) \cdot \sum_{t=1}^{n}[\Lambda_t(\boldsymbol{\beta}) - \Lambda_t(\boldsymbol{\beta}_0)]^2 = o_p(n), \end{aligned}$$

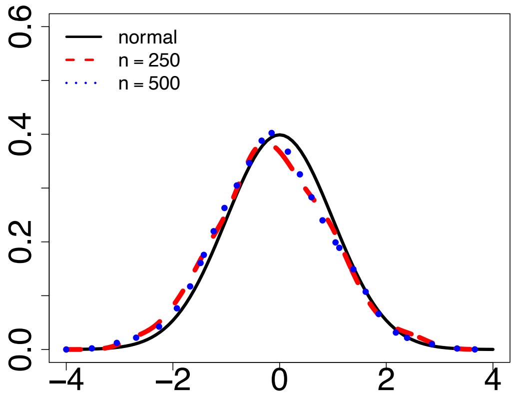
where $\lambda_{\max}(\boldsymbol{A})$ is the largest eigenvalue of matrix $\boldsymbol{A}$. In addition, we have $L_{n3} \leq (L_{n1}L_{n2})^{1/2} = o_p(1)$. Thus, the proof is complete.

D Additional Simulation Results

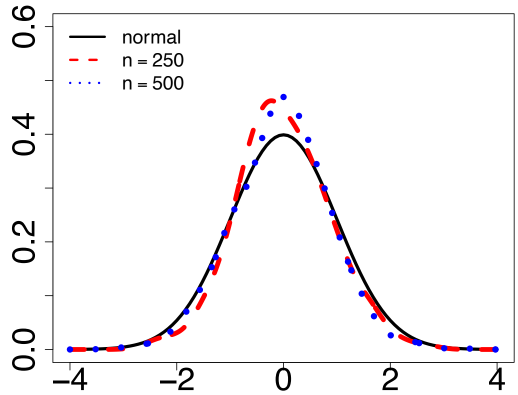
Some additional results are presented here. The findings are similar to those reported in the main paper, which therefore are not repeated.

Table 1: Bias, SD and RMSE ($\times 10^3$) for $\widehat{\beta}_n$ and $\widehat{\boldsymbol{\theta}}_n$ with $\rho = 0$, M1

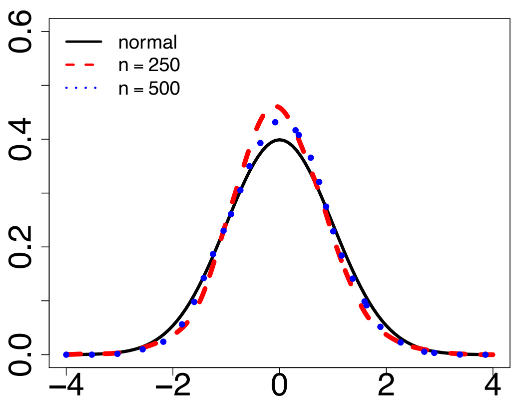
	n	250						500					
β_0		$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
							$d_t = 1$						
	BIAS	0.00	-0.22	-0.30	-0.53	-0.96	6.15	0.00	0.04	-0.12	-0.97	0.04	0.62
0.8	SD	0.34	7.31	7.52	44.2	44.8	121	0.15	3.56	3.65	30.9	32.5	88.1
	RMSE	0.34	7.32	7.53	44.2	44.8	121	0.15	3.56	3.65	30.9	32.5	88.1
	BIAS	-0.04	0.10	-0.47	-0.60	-2.27	1.23	0.00	-0.24	0.13	-0.29	0.87	-0.26
1	SD	0.40	7.18	7.63	45.1	44.6	117	0.17	3.64	3.76	32.0	30.5	82.8
	RMSE	0.40	7.18	7.64	45.1	44.7	117	0.17	3.64	3.76	32.0	30.5	82.8
	BIAS	-0.05	-0.03	-0.02	2.69	-1.68	-3.51	-0.01	-0.08	0.00	-0.16	0.47	0.98
1.2	SD	0.42	7.41	7.09	46.1	44.8	125	0.20	3.75	3.77	31.5	31.5	88.6
	RMSE	0.42	7.41	7.09	46.2	44.8	125	0.20	3.76	3.77	31.5	31.5	88.6
							$d_t = t$						
	BIAS	0.00	-0.08	0.02	-0.72	0.85	0.04	0.00	0.04	-0.17	1.60	0.47	0.01
0.8	SD	0.02	7.06	7.17	46.0	45.4	0.68	0.01	3.57	3.64	31.1	31.6	0.25
	RMSE	0.02	7.06	7.17	46.0	45.5	0.68	0.01	3.57	3.64	31.1	31.6	0.25
	BIAS	0.00	-0.19	0.29	-3.17	-0.24	-0.02	0.00	-0.24	0.11	-0.69	-0.26	0.01
1	SD	0.02	7.29	7.60	45.0	45.1	0.73	0.01	3.70	3.50	32.5	32.1	0.24
	RMSE	0.02	7.29	7.61	45.1	45.1	0.73	0.01	3.71	3.50	32.5	32.1	0.24
	BIAS	0.00	-0.14	0.07	-2.39	1.00	0.02	0.00	-0.06	0.12	0.06	-1.08	-0.01
1.2	SD	0.02	7.56	7.04	44.9	45.8	0.71	0.01	3.72	3.59	31.5	32.1	0.25
	RMSE	0.02	7.56	7.04	45.0	45.8	0.71	0.01	3.72	3.59	31.5	32.2	0.25



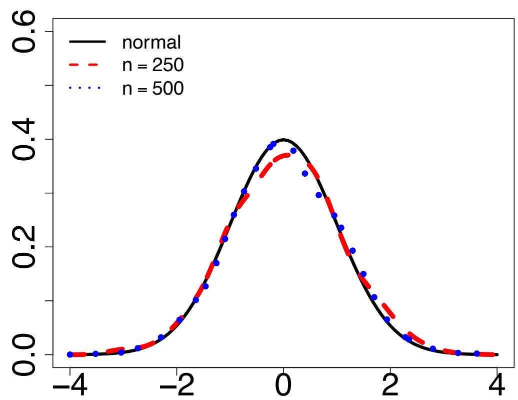
(a) $\widehat{\beta}_n$



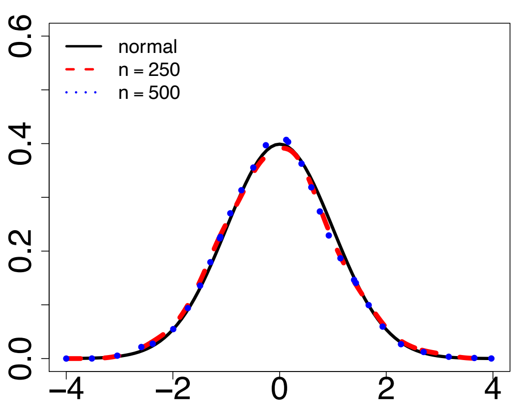
(b) $\widehat{\theta}_{n,11}$



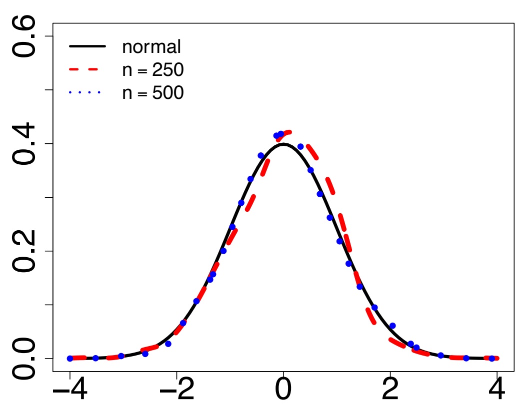
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$

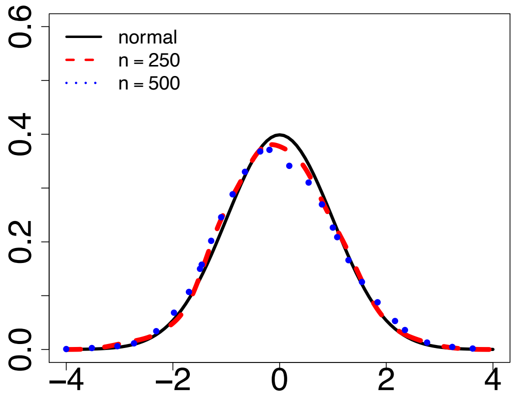


(e) $\widehat{\theta}_{n,22}$

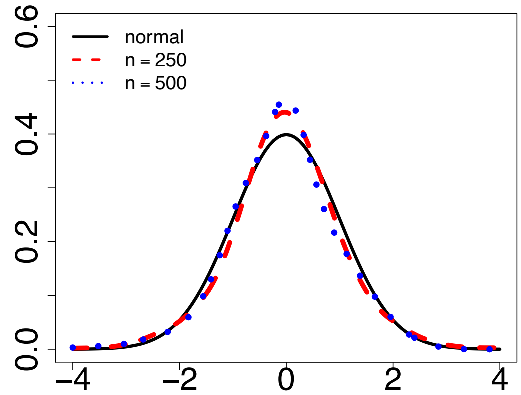


(f) $\widehat{\theta}_{n,3}$

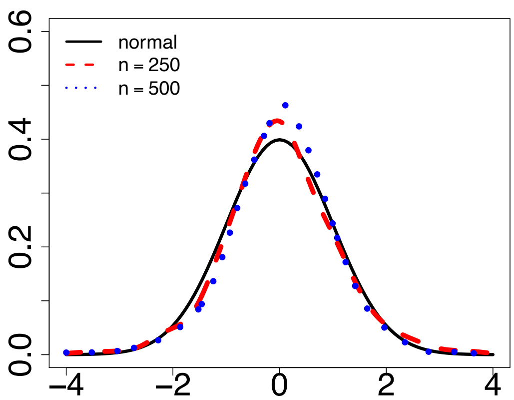
Figure D.1: Kernel density of t-ratios with $d_t = 1, \beta_0 = 0.8, \rho = 0.3$, M1.



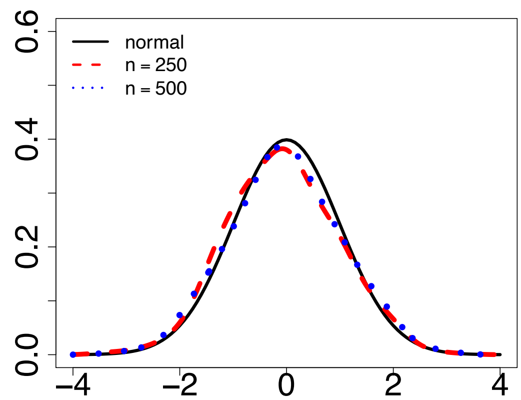
(a) $\widehat{\beta}_n$



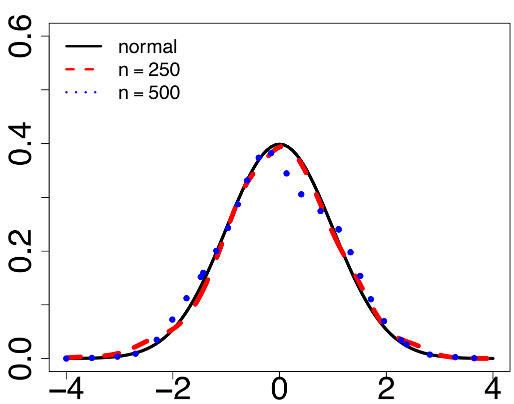
(b) $\widehat{\theta}_{n,11}$



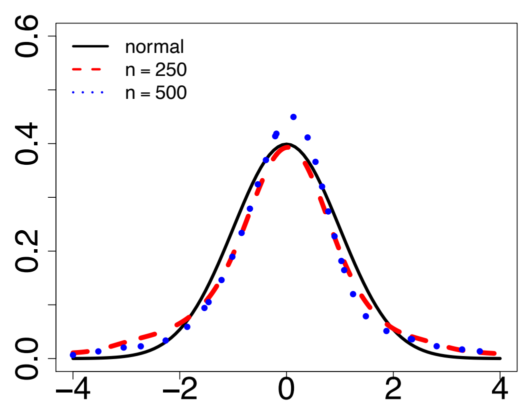
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$

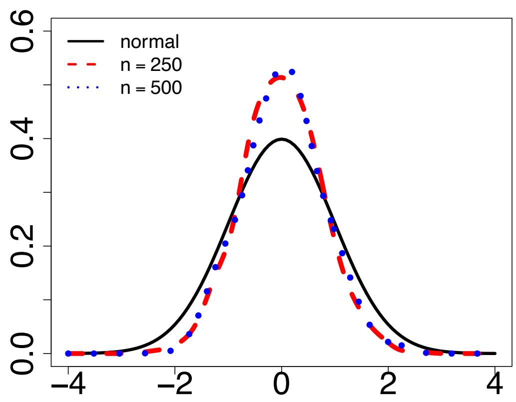


(e) $\widehat{\theta}_{n,22}$

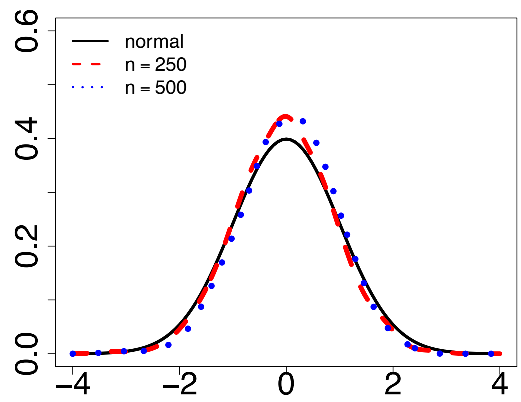


(f) $\widehat{\theta}_{n,3}$

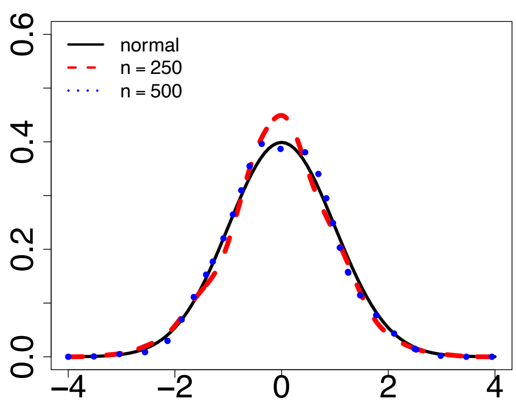
Figure D.2: Kernel density of t-ratios with $d_t = t, \beta_0 = 1, \rho = 0$, M1.



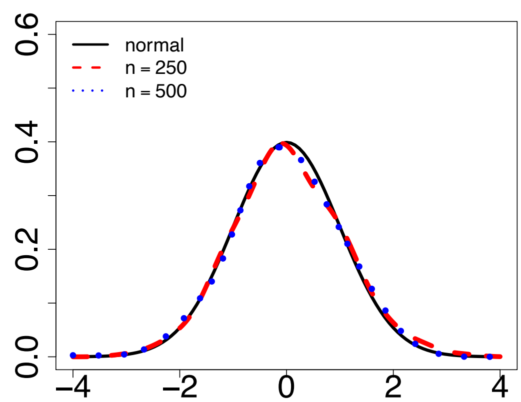
(a) $\widehat{\beta}_n$



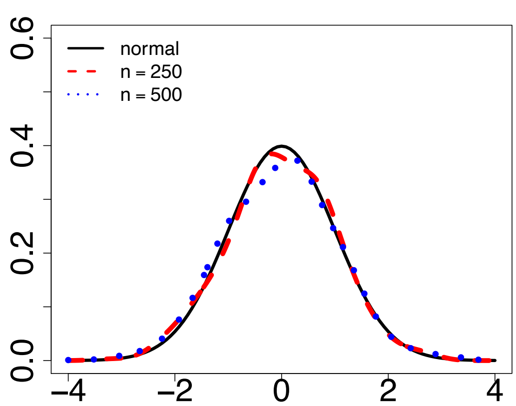
(b) $\widehat{\theta}_{n,11}$



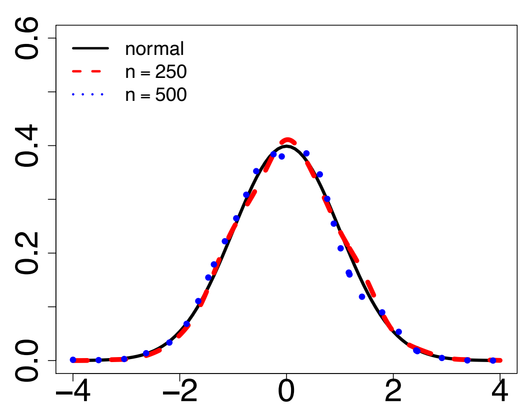
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$

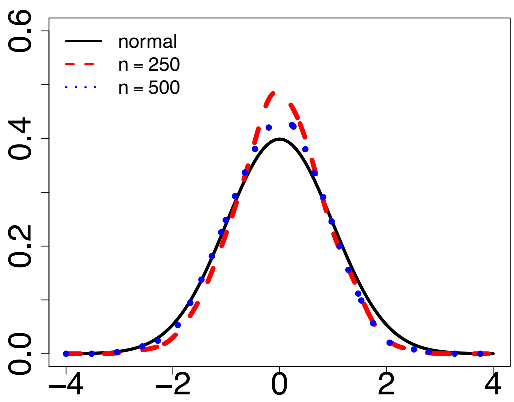


(e) $\widehat{\theta}_{n,22}$

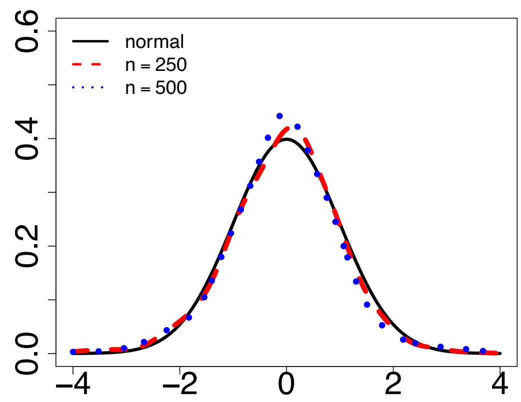


(f) $\widehat{\theta}_{n,3}$

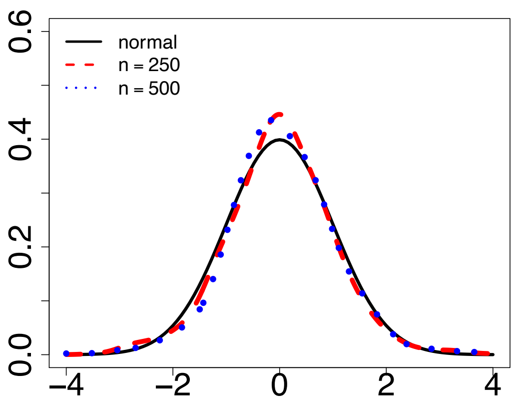
Figure D.3: Kernel density of t-ratios with $d_t = 1, \beta_0 = 1.2, \rho = 0.3$, M2.



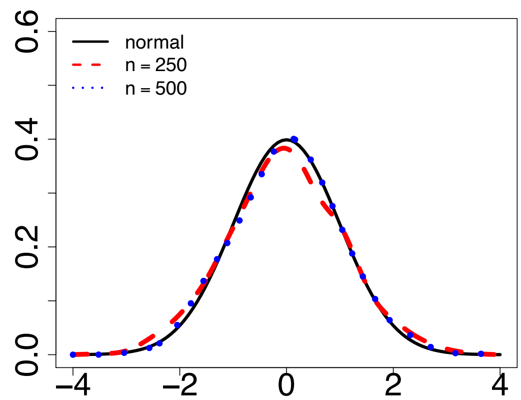
(a) $\widehat{\beta}_n$



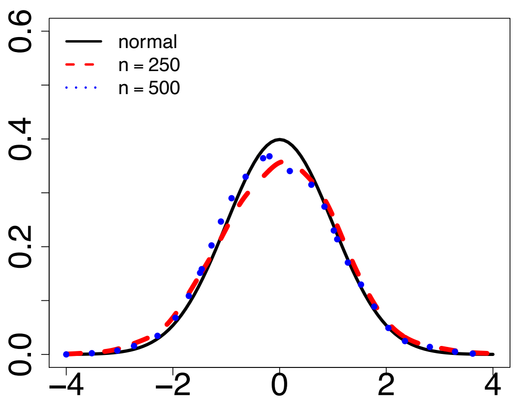
(b) $\widehat{\theta}_{n,11}$

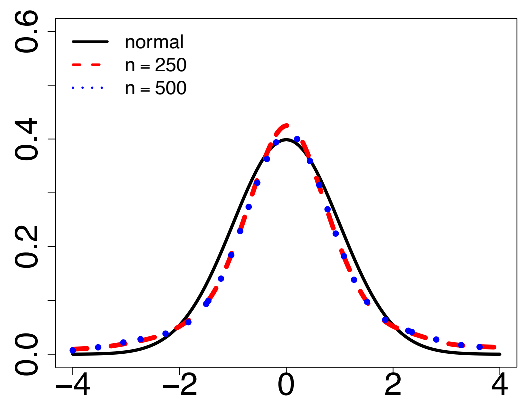


(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$

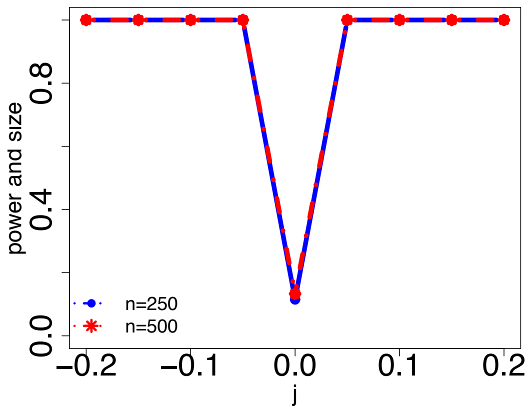




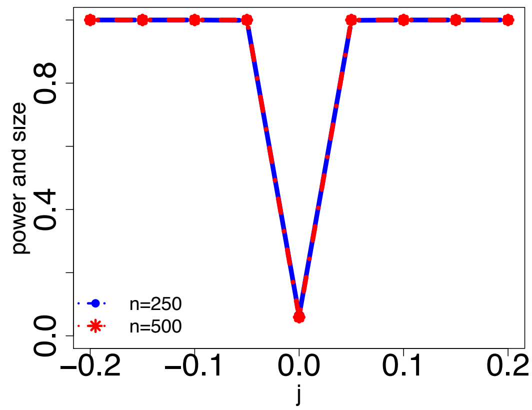
(e) $\widehat{\theta}_{n,22}$

(f) $\widehat{\theta}_{n,3}$

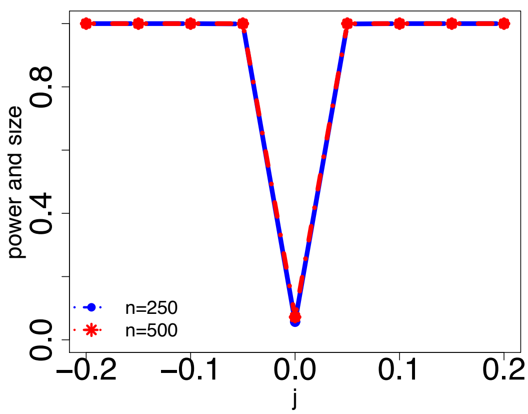
Figure D.4: Kernel density of t-ratios with $d_t = t, \beta_0 = 1, \rho = 0.7$, M2.



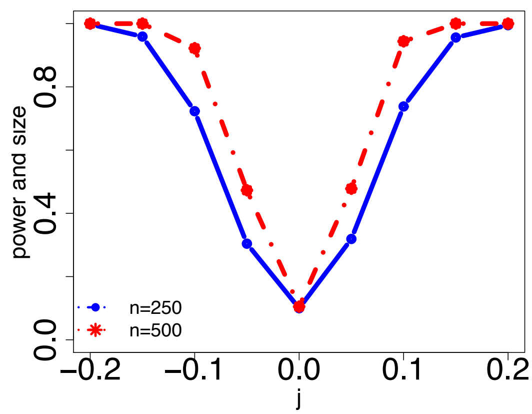
(a) $\widehat{\beta}_n$



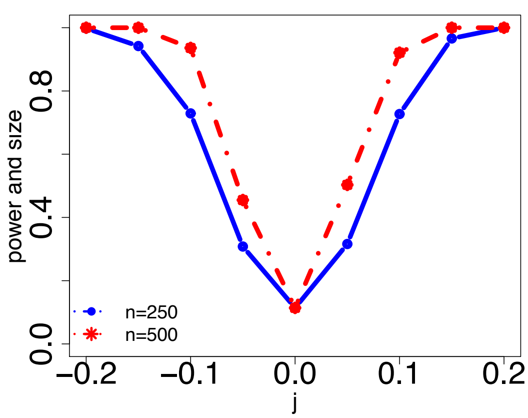
(b) $\widehat{\theta}_{n,11}$



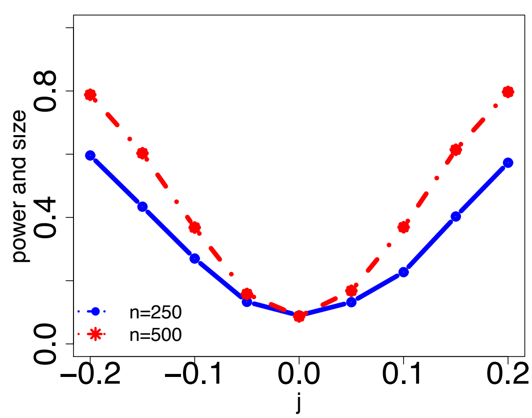
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$

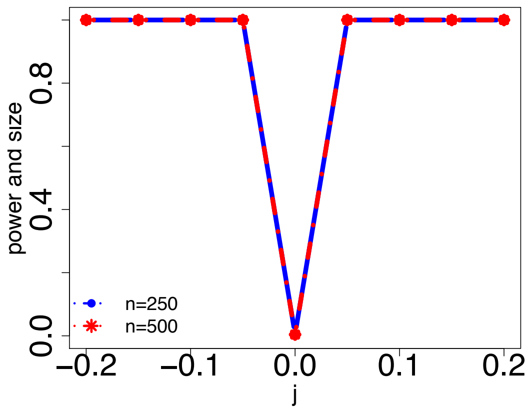


(e) $\widehat{\theta}_{n,22}$

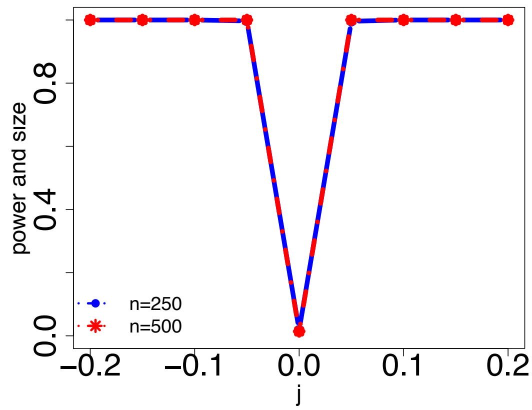


(f) $\widehat{\theta}_{n,3}$

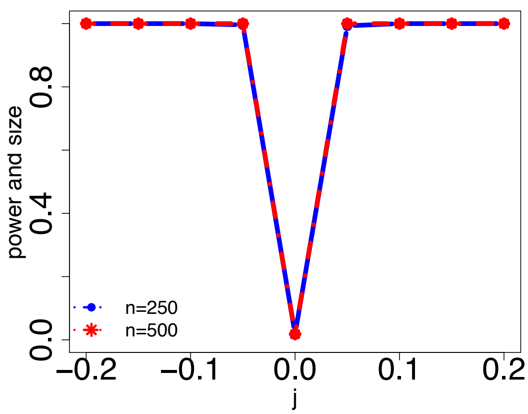
Figure D.5: Power and size ($j = 0$) of t-ratios with $\alpha = 0.10, d_t = 1, \beta_0 = 0.8, \rho = 0$, M1.



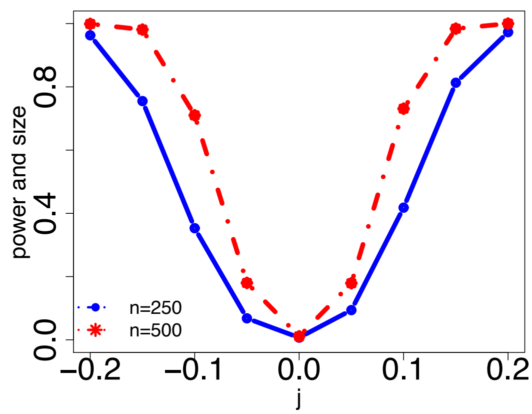
(a) $\widehat{\beta}_n$



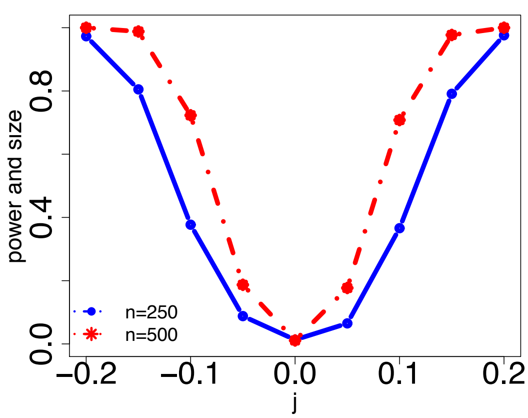
(b) $\widehat{\theta}_{n,11}$



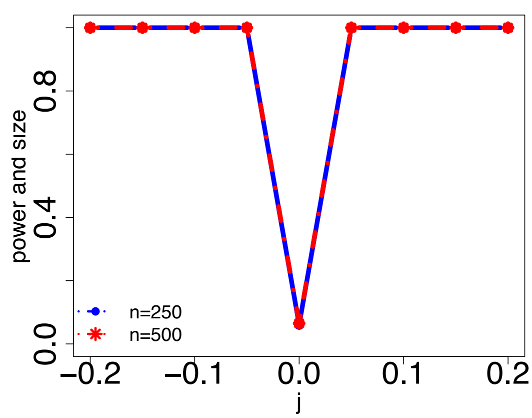
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$

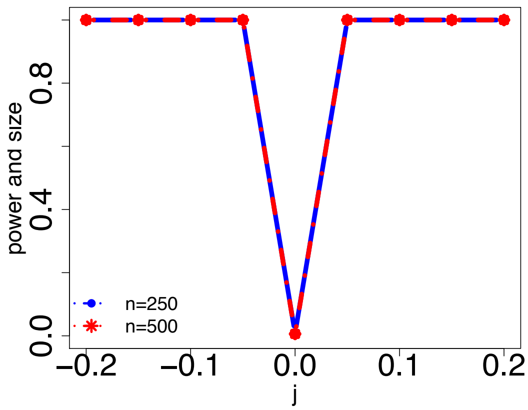


(e) $\widehat{\theta}_{n,22}$

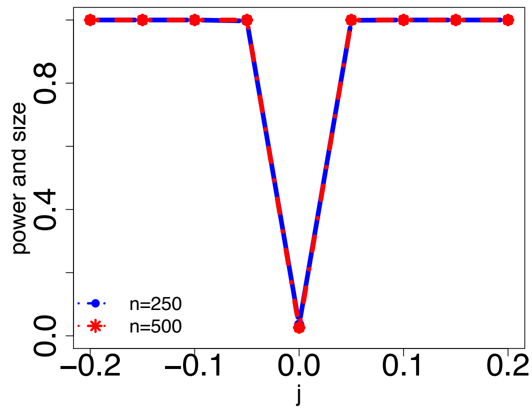


(f) $\widehat{\theta}_{n,3}$

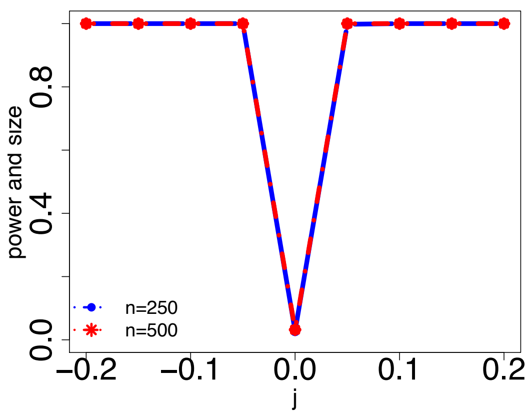
Figure D.6: Power and size ($j = 0$) of t-ratios with $\alpha = 0.01, d_t = t, \beta_0 = 1.2, \rho = 0.3$, M1.



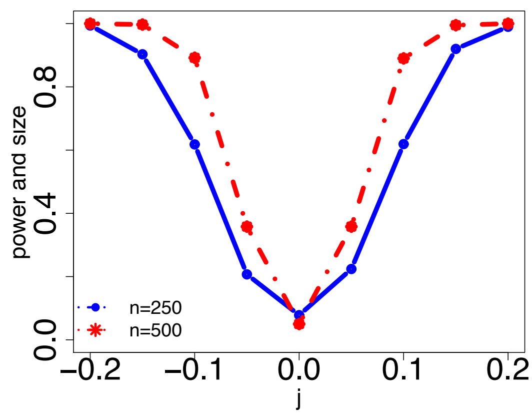
(a) $\widehat{\beta}_n$



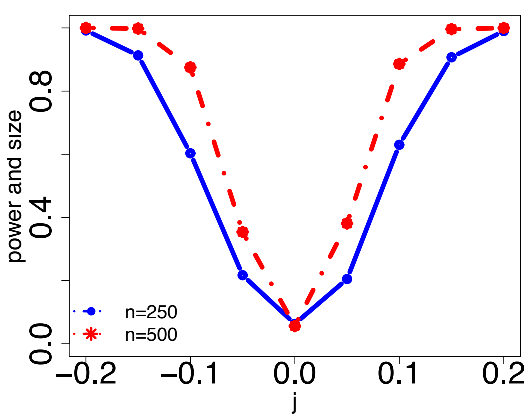
(b) $\widehat{\theta}_{n,11}$



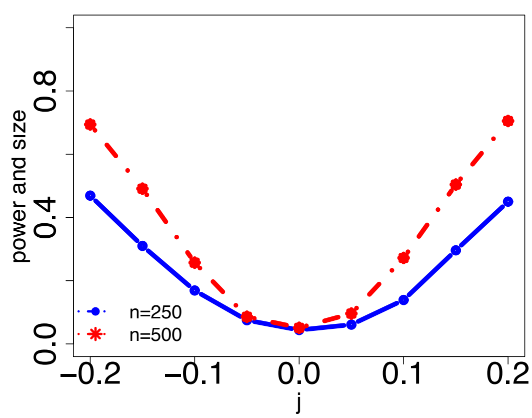
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$

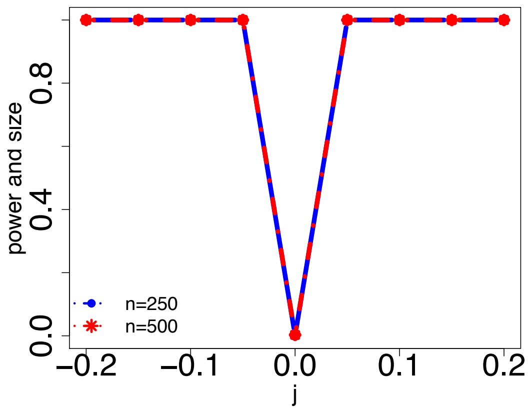


(e) $\widehat{\theta}_{n,22}$

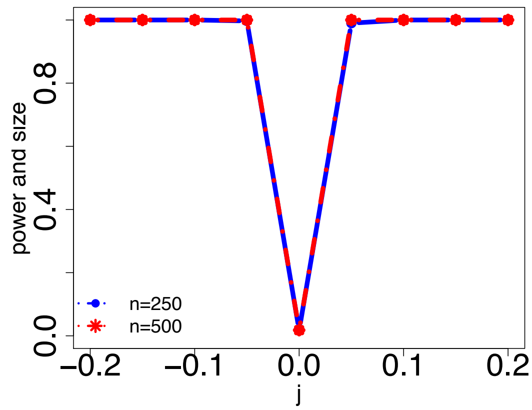


(f) $\widehat{\theta}_{n,3}$

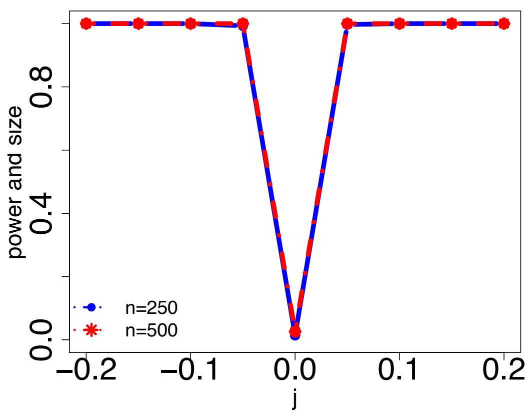
Figure D.7: Power and size ($j = 0$) of t-ratios with $\alpha = 0.05, d_t = 1, \beta_0 = 1, \rho = 0.3$, M2.



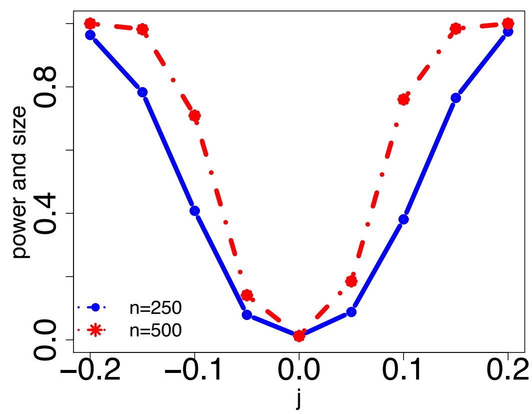
(a) $\widehat{\beta}_n$



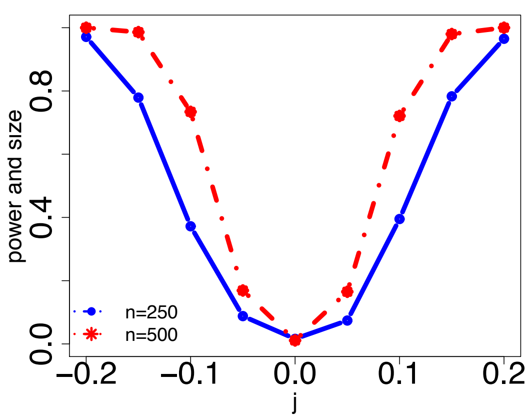
(b) $\widehat{\theta}_{n,11}$



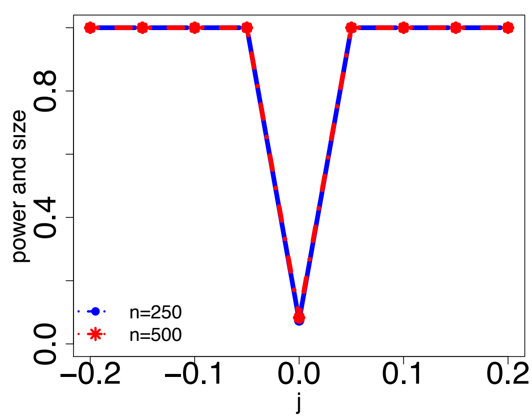
(c) $\widehat{\theta}_{n,12}$



(d) $\widehat{\theta}_{n,21}$



(e) $\widehat{\theta}_{n,22}$



(f) $\widehat{\theta}_{n,3}$

Figure D.8: Power and size ($j = 0$) of t-ratios with $\alpha = 0.01, d_t = t, \beta_0 = 1.2, \rho = 0$, M2.

Table 2: Ratios of the root mean squared errors for $n = 250$ to those for $n = 500$ with $\rho = 0$, M1

		β_0	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.225	2.0528	2.062	1.428	1.379	1.376
		1	2.350	1.969	2.031	1.411	1.465	1.419
		1.2	2.112	1.973	1.880	1.465	1.422	1.408
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	3.074	1.980	1.968	1.480	1.439	2.685
		1	3.052	1.967	2.175	1.385	1.408	3.077
		1.2	3.035	2.033	1.961	1.429	1.425	2.806

References

Chan, N. and Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, 185(1):182–195.

Chang, Y., Park, J. Y., and Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal*, 4(1):1–36.

Hu, Z., Phillips, P. C. B., and Wang, Q. (2021). Nonlinear cointegrating power function regression with endogeneity. *Econometric Theory*, doi:10.1017/S0266466620000560.

Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, 19(3):1035–1070.

Park, J. Y. and Phillips, P. C. B. (2000). Nonstationary binary choice. *Econometrica*, 68(5):1249–1280.

Table 3: Bias, SD and RMSE ($\times 10^3$) for $\widehat{\beta}_n$ and $\widehat{\boldsymbol{\theta}}_n$ with $\rho = 0.7$, M1

	n	250						500					
β_0		$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
							$d_t = 1$						
	BIAS	-0.04	0.16	-0.21	-0.08	1.07	-1.35	-0.01	-0.11	-0.21	0.73	-0.63	3.58
0.8	SD	0.34	7.38	7.63	45.4	46.2	124	0.15	3.54	3.68	31.6	31.8	87.2
	RMSE	0.34	7.38	7.64	45.4	46.2	124	0.15	3.55	3.69	31.6	31.8	87.3
	BIAS	-0.02	-0.21	-0.11	1.56	0.70	1.52	-0.01	-0.08	-0.18	0.50	1.58	2.56
1	SD	0.38	7.70	7.62	44.6	45.8	121	0.18	3.62	3.73	31.9	31.8	82.5
	RMSE	0.38	7.70	7.62	44.7	45.8	121	0.18	3.62	3.74	31.9	31.9	82.6
	BIAS	-0.04	-0.55	0.19	0.08	0.54	1.31	-0.01	-0.09	0.06	0.13	-0.43	1.06
1.2	SD	0.43	7.51	7.51	44.1	44.1	123	0.19	3.78	3.59	30.5	33.2	84.0
	RMSE	0.43	7.53	7.52	44.1	44.1	123	0.19	3.78	3.60	30.5	33.2	84.0
							$d_t = t$						
	BIAS	0.00	0.29	0.10	2.04	-2.12	-0.01	0.00	-0.10	0.04	1.22	-0.67	0.00
0.8	SD	0.02	7.65	7.12	44.8	44.8	0.69	0.01	3.71	3.46	32.8	31.4	0.25
	RMSE	0.02	7.65	7.12	44.8	44.8	0.69	0.01	3.71	3.46	32.9	31.4	0.25
	BIAS	0.00	0.25	-0.36	0.20	0.55	0.00	0.00	-0.09	0.04	1.24	-1.16	0.00
1	SD	0.02	8.04	7.48	44.5	46.8	0.72	0.01	3.65	3.59	31.6	31.5	0.25
	RMSE	0.02	8.05	7.48	44.5	46.8	0.72	0.01	3.66	3.59	31.7	31.5	0.25
	BIAS	0.00	0.11	0.10	1.78	1.08	-0.02	0.00	0.00	0.17	-0.54	-1.35	-0.01
1.2	SD	0.02	7.71	7.49	45.1	47.7	0.72	0.01	3.74	3.76	31.6	31.9	0.26
	RMSE	0.02	7.71	7.49	45.2	47.7	0.72	0.01	3.74	3.76	31.6	31.9	0.26

Table 4: Ratios of the root mean squared errors for $n = 250$ to those for $n = 500$ with $\rho = 0.7$, M1

		β_0	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.216	2.080	2.071	1.435	1.454	1.420
		1	2.161	2.125	2.041	1.400	1.438	1.461
		1.2	2.294	1.992	2.091	1.446	1.329	1.466
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	3.059	2.061	2.059	1.364	1.428	2.803
		1	2.987	2.201	2.085	1.406	1.485	2.878
		1.2	3.265	2.060	1.989	1.431	1.494	2.752

Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, 69(1):117–161.

Uematsu, Y. (2019). Nonstationary nonlinear quantile regression. *Econometric Reviews*, 38(4):386–416.

Table 5: Bias, SD and RMSE ($\times 10^3$) for $\widehat{\beta}_n$ and $\widehat{\boldsymbol{\theta}}_n$ with $\rho = 0.3$, M2

β_0	n	250						500					
		$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
		$d_t = 1$											
	BIAS	-0.02	0.05	-0.29	0.86	-0.24	0.63	-0.01	-0.07	-0.06	-0.09	0.03	2.40
0.8	SD	0.23	7.48	7.41	45.5	45.4	118	0.11	3.93	3.80	32.3	31.3	89.9
	RMSE	0.23	7.48	7.41	45.5	45.4	118	0.11	3.93	3.80	32.3	31.3	90.0
	BIAS	-0.03	-0.17	0.04	-1.52	-0.08	-1.61	-0.01	-0.01	-0.06	-0.09	-0.51	-2.26
1	SD	0.28	7.43	7.38	46.0	46.1	128	0.13	3.57	3.85	31.4	32.2	86.5
	RMSE	0.28	7.43	7.38	46.0	46.1	128	0.13	3.57	3.85	31.4	32.2	86.5
	BIAS	-0.01	-0.33	-0.29	1.64	-0.16	4.30	0.01	0.15	-0.14	-0.31	-0.96	-1.50
1.2	SD	0.36	7.55	7.72	45.2	43.8	121	0.16	3.61	3.93	31.6	32.7	90.4
	RMSE	0.36	7.56	7.73	45.2	43.8	121	0.16	3.61	3.93	31.6	32.7	90.4
		$d_t = t$											
	BIAS	0.00	0.59	-0.07	0.79	-2.00	-0.03	0.00	0.11	-0.16	-0.62	-0.22	0.00
0.8	SD	0.01	7.27	7.21	48.6	46.8	0.68	0.00	3.51	3.43	31.6	32.2	0.24
	RMSE	0.01	7.29	7.21	48.6	46.9	0.68	0.00	3.51	3.44	31.6	32.2	0.24
	BIAS	0.00	0.27	-0.40	-0.78	2.35	0.02	0.00	0.00	0.05	1.07	-2.34	0.00
1	SD	0.02	7.34	7.63	44.8	44.4	0.73	0.01	3.79	3.74	31.9	32.0	0.26
	RMSE	0.02	7.34	7.64	44.8	44.4	0.73	0.01	3.79	3.74	32.0	32.1	0.26
	BIAS	0.00	0.22	0.36	-3.68	1.20	-0.04	0.00	0.05	-0.32	-0.63	0.23	0.01
1	SD	0.02	7.18	7.16	43.9	43.5	0.68	0.01	3.50	3.66	32.2	32.6	0.24
	RMSE	0.02	7.18	7.16	44.1	43.5	0.68	0.01	3.50	3.67	32.2	32.6	0.25

Table 6: Ratios of the root mean squared errors for $n = 250$ to those for $n = 500$ with $\rho = 0.3$, M2

		β_0	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.076	1.902	1.948	1.411	1.452	1.307
		1	2.142	2.084	1.918	1.468	1.433	1.482
		1.2	2.277	2.094	1.964	1.428	1.338	1.343
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	2.864	2.079	2.097	1.536	1.457	2.853
		1	2.964	1.935	2.039	1.402	1.384	2.866
		1.2	3.195	2.052	1.951	1.368	1.336	2.773

Table 7: Bias, SD and RMSE ($\times 10^3$) for $\widehat{\beta}_n$ and $\widehat{\boldsymbol{\theta}}_n$ with $\rho = 0.7$, M2

	n	250						500					
β_0		$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
		$d_t = 1$											
	BIAS	-0.01	-0.06	0.32	0.22	-1.28	-7.14	0.00	0.03	-0.12	1.67	1.17	0.19
0.8	SD	0.24	7.66	7.84	45.6	47.2	122	0.11	4.03	3.64	31.7	31.7	89
	RMSE	0.24	7.66	7.84	45.6	47.2	123	0.11	4.03	3.64	31.8	31.7	89
	BIAS	-0.02	0.31	-0.09	1.62	-0.90	-4.49	-0.01	-0.01	-0.06	1.32	0.75	-0.32
1	SD	0.29	7.21	7.92	44.5	45.1	127	0.14	3.74	3.68	32.8	32.6	85
	RMSE	0.29	7.22	7.92	44.6	45.1	127	0.14	3.74	3.68	32.8	32.6	85
	BIAS	-0.02	0.00	-0.13	-2.04	-0.78	-2.25	-0.01	-0.08	0.02	-1.66	-0.64	-0.18
1.2	SD	0.34	6.98	7.43	45.0	42.3	115	0.16	3.80	3.89	31.4	32.7	89
	RMSE	0.34	6.98	7.44	45.1	42.3	115	0.16	3.80	3.89	31.4	32.7	89
		$d_t = t$											
	BIAS	0.00	0.33	-0.34	0.71	-0.38	0.01	0.00	0.14	-0.11	-0.51	0.28	-0.01
0.8	SD	0.01	7.55	7.34	44.7	44.2	0.77	0.00	3.65	3.60	32.7	31.4	0.23
	RMSE	0.01	7.56	7.35	44.7	44.2	0.77	0.00	3.65	3.60	32.7	31.4	0.23
	BIAS	0.00	-0.14	-0.08	-0.91	0.31	0.03	0.00	-0.18	0.10	0.52	-1.31	0.00
1	SD	0.02	7.41	7.63	46.7	47.6	0.69	0.01	3.82	3.69	31.1	32.4	0.27
	RMSE	0.02	7.41	7.63	46.7	47.6	0.69	0.01	3.83	3.69	31.1	32.4	0.27
	BIAS	0.00	0.38	0.49	-0.48	-1.95	-0.08	0.00	0.14	-0.13	-0.90	1.00	0.00
1.2	SD	0.02	7.09	7.77	45.0	45.8	0.72	0.01	3.55	3.61	31.9	32.1	0.26
	RMSE	0.02	7.10	7.79	45.0	45.8	0.73	0.01	3.55	3.62	32.0	32.1	0.26

Table 8: Ratios of the root mean squared errors for $n = 250$ to those for $n = 500$ with $\rho = 0.7$, M2

		β_0	$\widehat{\beta}_n$	$\widehat{\theta}_{n,11}$	$\widehat{\theta}_{n,12}$	$\widehat{\theta}_{n,21}$	$\widehat{\theta}_{n,22}$	$\widehat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.179	1.898	2.154	1.435	1.486	1.371
		1	2.157	1.929	2.149	1.360	1.385	1.497
		1.2	2.058	1.836	1.910	1.436	1.293	1.300
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	3.019	2.069	2.040	1.368	1.408	3.319
		1	2.969	1.936	2.066	1.503	1.469	2.545
		1.2	3.085	1.999	2.153	1.409	1.427	2.794