MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Parameter Stability Testing for Multivariate Dynamic Time-Varying Models

Jiti Gao, Bin Peng and Yayi Yan

October 2021

Working Paper 11/21

AACSB ACCREDITED EQUIS ACCREDITED AMBA ACCREDITED

Parameter Stability Testing for Multivariate Dynamic Time-Varying Models

Jiti Gao, Bin Peng and Yayi Yan

Monash University

September 5, 2021

Abstract

Multivariate dynamic models are widely used in practical studies providing a tractable way to capture evolving interrelationships among multivariate time series, but not many studies focus on inferences. Along this line, a key question is that whether some coefficients (if not all) evolve with time. To settle this issue, the paper develops a Wald–type test statistic for detecting time–invariant parameters in a class of multivariate dynamic time-varying models. Since Gaussian/stationary approximation methods initially proposed for univariate time series settings are inapplicable to the setting under consideration in this paper, we develop an approximation method using a time–varying vector moving average infinity process. We show that the test statistic is asymptotically normal under both the null hypothesis and the local alternative. Simulation studies show that the proposed test has a desirable finite sample performance.

Keywords: Multivariate Time Series; Parameter Instability; Specification testing; Time–varying coefficient

JEL classification: C12, C14, C32

Corresponding author: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Email: Jiti.Gao@monash.edu

1 Introduction

Multivariate dynamic models are widely used in various disciplines, e.g., psychological analysis (Haslbeck et al., 2021), network analysis (Geraci and Gnabo, 2018), monetary policy (Primiceri, 2005; Paul, 2020), etc. The use of multivariate dynamic models provides a tractable way to capture time-varying interrelationships among multivariate time series, and also allows us to mimic changes in unconditional volatility. However, it is difficult to find a consensus to represent multivariate dynamic models in practice, and more often than not, by employing different models, researches always lead to different conclusions. One classic example is the on-going debate on the high inflation during 1970-1980: bad policy or bad luck. We refer interested readers to Primiceri (2005) and Sims and Zha (2006) for more details on the debate. A similar concern has also been raised when studying the effects of monetary policy on asset prices. For example, Paul (2020) applies a constant VAR with exogenous variables (VARX) and a stochastic time-varying VARX model to investigate the effects of the monetary policy on asset prices, in which unsurprisingly certain inconsistency is observed. As a consequence, it is unclear which model produces more convincing results in a data driven fashion.

Mathematically, the aforementioned discrepancies come down to testing the coefficient matrices of multivariate dynamic models. Despite the importance, not much work has been done along this line of research. The literature to date focuses on detecting abrupt structural breaks in multivariate time series (e.g., Bai et al., 1998; Tsay, 1998; Bergamelli et al., 2019). Nevertheless, as pointed out by Hansen (2001), "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect". In addition, Primiceri (2005) discusses the advantages of a multivariate model with smooth time-varying coefficients from an economic perspective. That said, we aim to bridge this gap, i.e., providing hypothesis testing theory for multivariate dynamic models of which some (if not all) of the coefficient matrices transit over time in a smooth manner.

This paper therefore develops a hypothesis testing theory for multivariate dynamic models under a nonparametric framework. Specifically, we work with nonparametric time-varying VARX models, in which an unknown subset of the coefficients are allowed to be time-invariant. We construct a Wald-type test based on the L_2 distance between a nonparametric estimator and a semiparametric estimator. The test provides statistical evidence for us to allow for a subset of time–varying parameters of interest to be a subset of time–invariant constants in empirical analysis. Consequently, the rate of convergence on the time-invariant components can be improved to reach a parametric rate, which is in line with the discussion of Kilian and Lütkepohl (2017, pp.334) on how to boost the estimation accuracy for a multivariate system with comparatively short samples.

We now comment on the literature related to the hypothesis testing from a technical perspective. Detecting and estimating a parametric component in univariate time-varying models has been studied by Zhang and Wu (2012). Key elements in the proofs of Zhang and Wu (2012) include a Gaussian approximation to partial sums of nonlinear locally stationary processes and an extreme value theory for Gaussian processes. More recently, Truquet (2017) considers parameter stability testing for time-varying autoregressive conditional heteroscedasticity, and establishes asymptotic results for nonlinear locally stationary processes based on a stationary approximation method. The local stationarity approximation method generates a stochastic bias term of order $O_P(h)$, however, which is far larger than the usual deterministic bias term $O(h^2)$ in the literature of nonparametric kernel estimation. As a consequence, in order to obtain a meaningful distribution, one has to assume $Th^{3.5} \to 0$ as $T \to \infty$ where T is the sample size. Not only there is no guide for such bandwidth choice, but also the estimates are too volatile in finite sample studies. In this paper, we adopt a different framework, which makes the best use of an explicit structure of the data generating process of the multivariate system. A key step is to establish an approximation using a time–varying vector moving average infinity process (Lemma S1 of the supplementary file). Specifically, we work with a class of time-varying vector linear processes, which includes a large class of multivariate dynamic models. Under some weak conditions, such as $Th^{5.5} \to 0$, we have established the main results about the test statistic in Theorems 1 and 2.

Before proceeding further, it is convenient to introduce some notation. For a matrix A, $\|A\|$ denotes the Frobenius norm. Let $\otimes$ denote the Kronecker product. I_a and 0_a are $a \times a$ identity and null matrices respectively, and $0_{a\times b}$ stands for a $a \times b$ matrix of zeros. For a function $g(w)$, let $g^{(j)}(w)$ be the jth derivative of $g(w)$ with $j \geq 0$ and $g^{(0)}(w) \equiv g(w)$; $K_h(\cdot) = K(\cdot/h)/h$, where $K(\cdot)$ and h stand for a nonparametric kernel function and a bandwidth, respectively. $\text{vec}(\cdot)$ stacks the elements of an $m \times n$ matrix as an $mn \times 1$ vector. Finally, let $\to_D$ denote convergence in distribution, respectively.

2 Methodology

2.1 The setup

Having presented the motivation in Section 1, we specifically consider the following model:

$$y_t = \sum_{j=1}^{p} A_j(\tau_t) y_{t-j} + \sum_{j=0}^{q} B_j(\tau_t) x_{t-j} + \eta_t = Z_t^\top \beta(\tau_t) + \eta_t, \tag{2.1}$$

where $\tau_t = t/T$ with $t = 1, \ldots, T$, $Z_t = z_t \otimes I_d$, $z_t = (y_{t-1}^\top, \ldots, y_{t-p}^\top, x_t^\top, x_{t-1}^\top, \ldots, x_{t-q}^\top)^\top$, and $\eta_t = \omega(\tau_t)\epsilon_t$. Here, $y_t = (y_{1,t}, \ldots, y_{d,t})^\top$ is a d-dimensional vector of endogenous variables, $x_t = (x_{1,t}, \ldots, x_{m,t})^\top$ is an m-dimensional vector of exogenous variables, and both d and m are positive finite integers. Accordingly, $\{A_j(\tau)\}$ and $\{B_j(\tau)\}$ are the $d \times d$ and $d \times m$ coefficient matrices. Also, $\omega(\tau)$ is an unknown deterministic function which has full row rank uniformly in $\tau \in [0, 1]$, and captures the unconditional heteroscedasticity over time. Obviously, we have $\beta(\tau) = \text{vec}\{A(\tau), B(\tau)\}$, where $A(\tau) = (A_1(\tau), \ldots, A_p(\tau))$ and $B(\tau) = (B_0(\tau), B_1(\tau), \ldots, B_q(\tau))$.

In what follows, we infer $\{A_j(\tau)\}$ and $\{B_j(\tau)\}$, and are particularly interested in knowing whether some components of the coefficient matrices are constant. Mathematically, it is formulated as follows:

$$\mathbb{H}_0 : C\beta(\cdot) = c \text{ for some unknown } c \in \mathbb{R}^s, \tag{2.2}$$

where C is a selection matrix. Practically, the choice of C should be theory/application driven, and c needs to be estimated. In the case of monetary policy analysis (Paul, 2020), one can choose $C = (I_{d^2p}, 0_{d^2p\times(q+1)md})$ to test whether the policy transmission mechanism is varying over time. In addition, one can set C to be an identity matrix for selecting between stationary and time-varying VARX models. If a stationary VARX model is accepted, one can further consider the zero-nonzero pattern or general linear restrictions of $A_j(\cdot)$ and $B_j(\cdot)$ (cf., Guo et al., 2016; Zheng and Cheng, 2021).

The test statistic is constructed based on the weighted integrated squared errors

$$\widehat{Q}_{C,H} = \int_0^1 \left\{C\widehat{\beta}(\tau) - \widehat{c}\right\}^\top H(\tau) \left\{C\widehat{\beta}(\tau) - \widehat{c}\right\} \mathrm{d}\tau, \tag{2.3}$$

where $\widehat{\beta}(\cdot)$ and $\widehat{c}$ denote the estimators of $\beta(\cdot)$ and c respectively, and are to be constructed in Section 2.2. In (2.3), $H(\cdot)$ is an $s \times s$ positive definite weighting matrix, and is typically set as the precision matrix associated with $\widehat{\beta}(\cdot)$.

Finally, we emphasize that in order to analyze (2.3), a key step is to find its approximation using a time–varying vector moving average infinity process (Lemma S1 of the supplementary file).

2.2 On estimation

We are now ready to discuss the estimation of $\beta(\cdot)$ and c, and particularly adopt the nonparametric local linear approach and the profile likelihood estimation of Fan and Huang (2005). We start with $\beta(\cdot)$, and assume the components of $\beta(\cdot)$ have continuous derivatives up to the second order. When τ_t is close to τ, we then have the following approximation:

$$y_t \simeq Z_t^\top \beta(\tau) + Z_t^\top \beta^{(1)}(\tau)(\tau_t - \tau) + \eta_t. \tag{2.4}$$

Usually, $\{\beta(\tau), \beta^{(1)}(\tau)\}$ can be estimated by minimising the kernel weighted least-squares criterion:

$$(\widehat{\beta}(\tau), \widehat{\beta}^{(1)}(\tau)) = \underset{b_0,b_1}{\operatorname{argmin}} \sum_{t=1}^{T} \left\| y_t - Z_t^\top \{b_0 + (\tau_t - \tau)b_1\} \right\|^2 K_h(\tau_t - \tau). \tag{2.5}$$

Moreover, $\widehat{\beta}(\tau)$ admits a closed-form expression as follows:

$$\widehat{\beta}(\tau) = (I_l, 0_l)(Z_\tau^\top K_\tau Z_\tau)^{-1} Z_\tau^\top K_\tau y, \tag{2.6}$$

where $l = d^2p + (q+1)md$, $y = (y_1^\top, \dots, y_T^\top)^\top$, $K_\tau = \text{diag}\{K_h(\tau_1 - \tau), \dots, K_h(\tau_T - \tau)\} \otimes I_d$, and

$$Z_\tau = \begin{pmatrix} Z_1^\top & Z_1^\top \frac{\tau_1 - \tau}{h} \\ \vdots & \vdots \\ Z_T^\top & Z_T^\top \frac{\tau_T - \tau}{h} \end{pmatrix}.$$

Next, we estimate c. Under the null, we have $C\beta(\tau) = c$. To facilitate the development, we let $\widetilde{C}$ be a selection matrix collecting the elements of $\beta(\tau)$ left out by C. Thus, (2.1) can be rewritten as:

$$y_t = X_{C,t}^\top c + X_{\widetilde{C},t}^\top \theta(\tau_t) + \eta_t, \tag{2.7}$$

where $X_{C,t} = CZ_t$, $X_{\widetilde{C},t} = \widetilde{C}Z_t$, and $\theta(\tau) = \widetilde{C}\beta(\tau)$. The right hand side of (2.7) reduces to a semiparametric partially time-varying model. If the rank of C is l, then under the null model (2.1) reduces to a parametric model. By using the profile likelihood method to eliminate the nonparametric component $\theta(\tau)$, c can be estimated by

$$\widehat{c} = \left(X_C^\top (I_{dT} - S)^\top (I_{dT} - S) X_C\right)^{-1} X_C^\top (I_{dT} - S)^\top (I_{dT} - S) y, \tag{2.8}$$

where $X_C = (X_{C,1}, \dots, X_{C,T})^\top$ and $S = (s(\tau_1)^\top X_{\widetilde{C},1}, \dots, s(\tau_T)^\top X_{\widetilde{C},T})^\top$, in which

$$s(\tau) = (I_{l-s}, 0_{l-s})(X_{\widetilde{C},T\tau}^\top K_\tau X_{\widetilde{C},T\tau})^{-1} X_{\widetilde{C},T\tau}^\top K_\tau \quad \text{and} \quad X_{\widetilde{C},T\tau} = \begin{pmatrix} X_{\widetilde{C},1}^\top & X_{\widetilde{C},1}^\top \frac{\tau_1 - \tau}{h} \\ \vdots & \vdots \\ X_{\widetilde{C},T}^\top & X_{\widetilde{C},T}^\top \frac{\tau_T - \tau}{h} \end{pmatrix}.$$

2.3 On test statistic

We are now ready to study the test statistic (2.3) and investigate the local alternatives of (2.2). To facilitate the development, the following assumptions are necessary for the theoretical development.

Assumption 1. The roots of $I_d - A_1(\tau)L - \cdots - A_p(\tau)L^p = 0_d$ all lie outside the unit circle uniformly in $\tau \in [0,1]$.

Assumption 2. Each element of $\beta(\tau)$ is second-order continuously differentiable on $[0,1]$ and $\beta(\tau) = \beta(0)$ for $\tau < 0$.

Assumption 3. Suppose that $x_t = g(\tau_t) + \sum_{j=0}^{\infty} D_j(\tau_t) v_{t-j}$, where $g(\cdot)$ and $D_j(\cdot)$ are $m \times 1$ and $m \times m$ respectively. Each component of $g(\cdot)$ and $D_j(\cdot)$ is second-order continuously differentiable on $[0,1]$, and $g(\tau) = g(0)$ and $D_j(\tau) = D_j(0)$ for $\tau < 0$. For $\ell = 0,1$, $\sup_{\tau \in [0,1]} \sum_{j=1}^{\infty} j \|D_j^{(\ell)}(\tau)\| < \infty$.

Assumption 4. Each component of $\omega(\tau)$ is second-order continuously differentiable on $[0,1]$. Moreover, $\Omega(\tau) = \omega(\tau)\omega(\tau)^\top$ is positive definite uniformly in $\tau \in [0,1]$, and $\omega(\tau) = \omega(0)$ for $\tau < 0$.

Assumption 5. Let $e_t = (\epsilon_t^\top, v_{t+1}^\top)^\top$, and $\{e_t\}_{t=-\infty}^{\infty}$ be a sequence of martingale differences, that is $E\left(e_t \mid \mathcal{F}_{t-1}\right)=0$, where $\mathcal{F}_t=\sigma\{e_t, e_{t-1}, \ldots\}$. Also, suppose that

$$E(e_t e_t^{\top} \mid \mathcal{F}_{t-1})=\begin{pmatrix} I_d & \rho_{\epsilon v} \\ \rho_{\epsilon v}^{\top} & I_m \end{pmatrix}$$

almost surely, and $\max_{t\geq 1} E\left\|e_t\right\|^{\delta}<\infty$ for some $\delta>2$.

Assumption 6. Let $K(\cdot)$ be a symmetric kernel function that is Lipschitz continuous on $[-1,1]$, and satisfies $\int_{-1}^{1} K(u)\mathrm{d}u=1$. Also, let $h \to 0$ and $Th \to \infty$ as $T \to \infty$.

Assumption 7. Let $\max_{t\geq 1} E[\|e_t\|^{\delta} \mid \mathcal{F}_{t-1}]<\infty$ almost surely, $Th^8 \to 0$, $Th^2/(\log T)^2 \to \infty$, $(T^{1-4/\delta}h)/\log T \to \infty$, and $\delta>4$, where δ is the same as that of Assumption 5.

Assumptions 2 and 6 are pretty standard in the literature of nonparametric regression (Li and Racine, 2007), so the discussions are omitted. Assumption 5 is also standard by assuming that the innovation errors follow a martingale difference structure(Phillips and Lee, 2013).

We now comment on the rest of these conditions. Assumption 1 ensures that y_t in model (2.1) is neither a unit-root process nor an explosive process, and can be regarded as an extension of those used for classical multivariate dynamic models (e.g., Hamilton, 1994, p. 259). Assumption 3 formulates a time-varying vector moving average process (TV-VMA(∞)) which nests many commonly used processes as special cases. Extensive investigation on TV-VMA(∞) with examples can be seen in Yan et al. (2020), so we no longer expand the discussion here. Assumption 4 models the heteroscedasticity using an unknown function, and the conditions are in the same spirit as those for $\boldsymbol{\beta}(\cdot)$.

Assumption 7 imposes extra conditions on the bandwidth and the conditional moments of the error terms, which are commonly used in the literature of semiparametric regression (e.g., Fan and Huang, 2005; Chen et al., 2012).

Using Assumptions 1–7 in hand, after some tedious development, one actually can show that

$$\widehat{Q}_{C,H}=\frac{1}{Th}\cdot\frac{\widetilde{v}_0}{T}\sum_{t=1}^{T}\eta_t^{\top} Z_t^{\top} H_0(\tau_t) Z_t \eta_t+O_P\left(\frac{1}{T\sqrt{h}}\right), \tag{2.9}$$

where $\tilde{v}_0=\int_{-1}^{1} K^2(u)\mathrm{d}u$ and $H_0(\tau)$ is a weighting matrix whose definition is omitted here. Clearly, $\widetilde{v}_0 T^{-1}\sum_{t=1}^{T}\eta_t^{\top} Z_t^{\top} H_0(\tau_t) Z_t \eta_t$ converges to a fixed value in probability, so the first term on the right hand side of (2.9) is the leading one. Moreover, $\widetilde{v}_0 T^{-1}\sum_{t=1}^{T}\eta_t^{\top} Z_t^{\top} H_0(\tau_t) Z_t \eta_t$ can be considered as a weighted average of the second moment of η_t. It however does not yield any distribution for us to conduct the hypothesis testing. In this regard, it can be considered a bias term which should be removed. As a consequence, the term which yields the asymptotic distribution is in fact the one having order $O_P\left(\frac{1}{T\sqrt{h}}\right)$.

Below, we establish an asymptotic distribution for the test statistic of (2.3). Let $V(\tau)$ be the asymptotic covariance matrix of $\widehat{\beta}(\tau)$. Hence, one can choose $H(\tau)=(CV(\tau)C^{\top})^{-1}$ to serve as a normalizer. Then $\widehat{Q}_{C,H}$ is a pivotal test statistic. As $V(\tau)$ is unknown in general, we use its approximation which is explicitly defined in the next theorem.

THEOREM 1. *Suppose that Assumptions 1–7 hold. Under the null, if* $Th^{11/2} \to 0$,

$$T\sqrt{h}\left(\widehat{Q}_{C,\widehat{H}} - \frac{1}{Th}s\widetilde{v}_0\right) \to_D N(0, 4sC_B),$$

where $C_B = \int_0^2\{\int_{-1}^{1-v} K(u)K(u+v)\mathrm{d}u\}^2\mathrm{d}v$, $\widehat{H}(\tau) = (C\widehat{V}(\tau)C^\top)^{-1}$, $\widehat{V}(\tau) = \widehat{\Sigma}_z^{-1}(\tau) \otimes \widehat{\Omega}(\tau)$, $\widehat{\Sigma}_z(\tau) = \{\sum_{t=1}^T K_h(\tau_t - \tau)\}^{-1}\sum_{t=1}^T z_t z_t^\top K_h(\tau_t - \tau)$, $\widehat{\Omega}(\tau) = \{\sum_{t=1}^T K_h(\tau_t - \tau)\}^{-1}\sum_{t=1}^T \widehat{\eta}_t\widehat{\eta}_t^\top K_h(\tau_t - \tau)$, *and* $\widehat{\eta}_t = y_t - Z_t^\top\widehat{\beta}(\tau_t)$.

Theorem 1 states that the test statistic converges to a normal distribution. The term $s\widetilde{v}_0$ is the limit associated with $\widetilde{v}_0 T^{-1}\sum_{t=1}^T \eta_t^\top Z_t^\top H_0(\tau_t) Z_t \eta_t$ of (2.9). Due to the use of $\widehat{H}(\cdot)$, the second moment of η_t disappears from the asymptotic distribution automatically. Here, we would like to emphasize that the proposed test is in fact a one–sided test, which is reflected in the investigation for the local alternative below. An intuitive explanation is that due to the quadratic form of (2.3), any departure from the true value will eventually yield a squared term when analyzing the asymptotic power. That said, we conclude that the null of (2.2) will be rejected at the level α if

$$\widehat{Q}^*_{C,\widehat{H}} = \frac{T\sqrt{h}\left(\widehat{Q}_{C,\widehat{H}} - \frac{1}{Th}s\widetilde{v}_0\right)}{\sqrt{(4sC_B)}} > q_{1-\alpha}, \tag{2.10}$$

where $q_{1-\alpha}$ stands for the $(1-\alpha)$th quantile of the standard normal distribution.

In what follows, we consider the local alternative of the null in (2.2), and are specifically interested in the following form:

$$\mathbb{H}_1 : C\beta(\tau) = c + d_T f(\tau), \tag{2.11}$$

where $f(\tau)$ is a twice continuously differentiable vector function, and $d_T \to 0$ as $T \to \infty$. The term $d_T f(\tau)$ characterizes the departure of the time-varying coefficient $C\beta(\tau)$ from the constant c. Following the development of Theorem 1, it is straightforward to obtain the following theorem.

THEOREM 2. *Let the conditions of Theorem 1 hold. If* $d_T = T^{-1/2}h^{-1/4}$, *then under* $\mathbb{H}_1$ *of* (2.11)

$$T\sqrt{h}\left(\widehat{Q}_{C,\widehat{H}} - \frac{1}{Th}s\widetilde{v}_0\right) \to_D N\left(\delta_1, 4sC_B\right),$$

where $\delta_1 = \int_0^1 f(\tau)^\top(CV(\tau)C^\top)^{-1}f(\tau)\mathrm{d}\tau$. *Moreover,*

$$\mathrm{pr}\left(\widehat{Q}^*_{C,\widehat{H}} > q_{1-\alpha}\right) \to \Phi\left(q_\alpha + \frac{\delta_1}{\sqrt{(4sC_B)}}\right),$$

where $\Phi(\cdot)$ *is the CDF of the standard normal distribution.*

Theorem 2 shows that the test has a non-trivial power against $\mathbb{H}_1$ with $d_T = T^{-1/2}h^{-1/4}$. If $T^{-1/2}h^{-1/4} = o(d_T)$, the power of the test converges to 1, i.e., $\mathrm{pr}(\widehat{Q}^*_{C,\widehat{H}} > q_{1-\alpha}) \to 1$.

2.4 Implementation of the test

It is known that test statistics associated with asymptotic critical values may not have good finite–sample sizes and power values. To improve the finite sample performance of our test, we propose a simulation-assisted testing procedure. The following algorithm outlines the steps to perform the test. The Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \le 1)$ is adopted throughout the numerical studies. For bandwidth selection, our simulation results in Section 3 show that the empirical size of our test is not sensitive to the choice of bandwidth. For simplicity, one can choose the rule of thumb bandwidth $h = 2.34\sqrt{(1/12)}T^{-1/5}$ that has the mean squared error optimal rate.

- **Algorithm: a simulation-assisted testing procedure**
  - Input: Time series $\{y_t, x_t\}$, bandwidth h and selection matrix C.

    Step 1. Use the sample $\{y_t, x_t\}$ to estimate the unrestricted model and the restricted model, and then compute $\widehat{Q}_{C,\widehat{H}}$ based on (2.3).

    Step 2. Generate i.i.d. standard multivariate normal random vectors $\{y_t^*\}$ and $\{x_t^*\}$ in which $\{y_t^*\}$ are independent on $\{x_t^*\}$.

    Step 3. Compute the bootstrap statistic $\widetilde{Q}^b_{C,\widehat{H}}$ in the same way as $\widehat{Q}_{C,\widehat{H}}$, with $\{y_t^*, x_t^*\}$ replacing the original sample $\{y_t, x_t\}$.

    Step 4. Repeat Steps 2–3 B times to obtain B bootstrap test statistics $\{\widetilde{Q}^b_{C,\widehat{H}}\}_{b=1}^B$, as well as its empirical quantile $\widehat{q}_{1-\alpha}$. We reject the null hypothesis (2.2) at level α if $\widehat{Q}_{C,\widehat{H}} > \widehat{q}_{1-\alpha}$.

A similar procedure has also been adopted by Zhang and Wu (2012) and Truquet (2017) to improve the finite sample performance in the context of univariate time-varying models.

3 Simulation

To evaluate the size and local power of the proposed test statistic, we consider the following data generating process:

$$y_t = A_1(\tau)y_{t-1} + B_1(\tau)x_{t-1} + \epsilon_t,$$

where ϵ_t's are i.i.d. draws from $N(0_{2\times 1}, I_2)$,

$$A_1(\tau) = \begin{bmatrix} 0.4 & -0.1 \\ -0.1 & 0.4 \end{bmatrix} + b \times d_T \times \begin{bmatrix} 2\exp(\tau-1)-1 & \exp(\tau-1)-1 \\ \exp(\tau-1)-1 & 2\exp(\tau-1)-1 \end{bmatrix},$$

$$B_1(\tau) = [2\exp(\tau-1)-1, 2\exp(\tau-1)-1]^\top,$$

in which $d_T = T^{-1/2}h^{-1/4}$ and b is set to be 0, 2 or 4 in order to investigate the size and local power of the proposed test. In addition, x_t is generated by $x_t = 0.4x_{t-1} + v_t$ with $v_t \sim$ i.i.d.$N(0, 1)$.

Table 1: Reject frequency for partially time–varying test.

	Bandwidth	T	5% Nominal Level 200	400	800	10% Nominal Level 200	400	800
$b = 0$ (size)	$0.2T^{-1/5}$		0.093	0.098	0.068	0.159	0.150	0.122
	$0.4T^{-1/5}$		0.084	0.050	0.057	0.141	0.100	0.098
	$0.6T^{-1/5}$		0.044	0.061	0.062	0.095	0.105	0.100
	$0.8T^{-1/5}$		0.049	0.027	0.041	0.084	0.077	0.082
	$1.0T^{-1/5}$		0.043	0.049	0.042	0.095	0.101	0.099
	$1.2T^{-1/5}$		0.051	0.054	0.050	0.099	0.108	0.089
	$1.4T^{-1/5}$		0.048	0.055	0.047	0.091	0.099	0.077
	$1.6T^{-1/5}$		0.046	0.050	0.056	0.089	0.093	0.124
	$1.8T^{-1/5}$		0.040	0.044	0.042	0.084	0.092	0.075
$b = 2$ (local power)	$0.2T^{-1/5}$		0.193	0.307	0.287	0.360	0.469	0.434
	$0.4T^{-1/5}$		0.235	0.268	0.261	0.347	0.397	0.386
	$0.6T^{-1/5}$		0.234	0.203	0.229	0.332	0.323	0.348
	$0.8T^{-1/5}$		0.198	0.203	0.239	0.338	0.337	0.353
	$1.0T^{-1/5}$		0.197	0.217	0.209	0.337	0.321	0.341
	$1.2T^{-1/5}$		0.165	0.177	0.209	0.290	0.277	0.294
	$1.4T^{-1/5}$		0.135	0.174	0.194	0.248	0.252	0.274
	$1.6T^{-1/5}$		0.194	0.160	0.159	0.300	0.256	0.277
	$1.8T^{-1/5}$		0.212	0.158	0.179	0.301	0.248	0.291
$b = 4$ (local power)	$0.2T^{-1/5}$		0.916	0.944	0.951	0.966	0.970	0.971
	$0.4T^{-1/5}$		0.898	0.923	0.914	0.942	0.961	0.954
	$0.6T^{-1/5}$		0.851	0.859	0.895	0.913	0.933	0.937
	$0.8T^{-1/5}$		0.779	0.799	0.857	0.865	0.886	0.925
	$1.0T^{-1/5}$		0.762	0.774	0.795	0.834	0.865	0.896
	$1.2T^{-1/5}$		0.735	0.770	0.803	0.812	0.850	0.859
	$1.4T^{-1/5}$		0.702	0.755	0.752	0.809	0.829	0.840
	$1.6T^{-1/5}$		0.672	0.700	0.754	0.770	0.796	0.847
	$1.8T^{-1/5}$		0.711	0.702	0.715	0.784	0.817	0.817

Specifically, we test

$$\mathbb{H}_0 : A_1(\cdot) \equiv A \text{ in which } A \text{ is an unknown constant matrix.}$$

We set T to be $200, 400, 800$ and conduct 1000 replications for each choice of T. For each setting, we use the bootstrap-assisted testing procedure to get the empirical critical value $\widehat{q}_{1-\alpha}$, and the number of bootstrap repetitions is 1000. As discussed in Gao and Gijbels (2008), estimation-based optimal bandwidths may still be optimal for testing purposes. We then consider the bandwidths $h = \alpha_1 T^{-1/5}$ with $\alpha_1 = 0.2, 0.4, \ldots, 1.8$.

Table 1 reports the rejection rates at the 5% and 10% nominal levels. A few facts emerge from Table 1. First, our test has reasonable sizes using empirical critical values obtained by the bootstrap procedure. Second, the size is not sensitive to the choice of bandwidths as long as the bandwidths are not so small. Third, the local power of the test increases rapidly as b increases. Moreover, the local power of the test increases, as the bandwidth h decreases. The above results indicate that the optimal bandwidth h for testing purposes is around $0.4T^{-1/5}$ to $0.8T^{-1/5}$, in which the local power is maximized and the size is very close to the nominal level. Since the range covers the rule of thumb bandwidth $h = 2.34\sqrt{(1/12)}T^{-1/5} = 0.676T^{-1/5}$, for simplicity the rule of thumb bandwidth might be a good choice in practice.

Acknowledgement

The authors of this paper would like to thank George Athanasopoulos, David Frazier, Gael Martin, Mervyn Silvapulle and Param Silvapulle for their constructive comments on earlier versions of this paper. Gao acknowledges financial support from the Australian Research Council Discovery Grants Program under Grant Numbers: DP170104421 and DP200102769. Peng also acknowledges the Australian Research Council Discovery Grants Program for its financial support under Grant Number DP210100476.

References

Bai, J., Lumsdaine, R. L. and Stock, J. H. (1998), 'Testing for and dating common breaks in multivariate time series', *Review of Economic Studies* **65**(3), 395–432.

Bergamelli, M., Bianchi, A., Khalaf, L. and Urga, G. (2019), 'Combining p-values to test for multiple structural breaks in cointegrated regressions', *Journal of Econometrics* **211**(2), 461–482.

Chen, J., Gao, J. and Li, D. (2012), 'Semiparametric trending panel data models with cross-sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Chen, Z. and Leng, C. (2015), 'Local linear estimation of covariance matrices via cholesky decomposition', *Statistica Sinica* **25**(3), 1249–1263.

Dahlhaus, R. and Polonik, W. (2009), 'Empirical spectral processes for locally stationary time series', *Bernoulli* **15**(1), 1–39.

Fan, J. and Huang, T. (2005), 'Profile likelihood inferences on semiparametric varying-coefficient partially linear models', *Bernoulli* **11**(6), 1031–1057.

Freedman, D. A. (1975), 'On tail probabilities for martingales', *Annals of Probability* **3**(1), 100–118.

Gao, J. and Gijbels, I. (2008), 'Bandwidth selection in nonparametric kernel testing', *Journal of the American Statistical Association* **103**(484), 1584–1594.

Geraci, M. V. and Gnabo, J.-Y. (2018), 'Measuring interconnectedness between financial institutions with Bayesian time-varying vector autoregressions', *Journal of Financial and Quantitative Analysis* **53**(3), 1371–1390.

Guo, S., Wang, Y. and Yao, Q. (2016), 'High-dimensional and banded vector autoregressions', *Biometrika* **103**(4), 889–903.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Academic Press.

Hamilton, J. D. (1994), *Time Series Analysis*, Vol. 2, Princeton University Press, New Jersey.

Hansen, B. E. (2001), 'The new econometrics of structural change: dating breaks in US labour productivity', *Journal of Economic Perspectives* **15**(4), 117–128.

Haslbeck, J. M., Bringmann, L. F. and Waldorp, L. J. (2021), 'A tutorial on estimating time-varying vector autoregressive models', *Multivariate Behavioral Research* **56**(1), 120–149.

Kilian, L. and Lütkepohl, H. (2017), *Structural Vector Autoregressive Analysis*, Cambridge University Press.

Li, Q. and Racine, J. (2007), *Nonparametric Econometrics Theory and Practice*, Princeton University Press, New Jersey.

Paul, P. (2020), 'The time–varying effect of monetary policy on asset prices', *Review of Economics and Statistics* **102**(4), 690–704.

Phillips, P. C. B. and Lee, J. H. (2013), 'Predictive regression under various degrees of persistence and robust long-horizon regression', *Journal of Econometrics* **177**(2), 250–264.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies* **72**(3), 821–852.

Sims, C. and Zha, T. (2006), 'Were there regime switches in U.S. monetary policy?', *American Economic Review* **96**(1), 54–81.

Truquet, L. (2017), 'Parameter stability and semiparametric inference in time varying auto-regressive conditional heteroscedasticity models', *Journal of the Royal Statistical Society: Series B* **79**(5), 1391–1414.

Tsay, R. S. (1998), 'Testing and modeling multivariate threshold models', *Journal of the American Statistical Association* **93**(443), 1188–1202.

Wu, W. B. and Shao, X. (2007), 'A limit theorem for quadratic forms and its applications', *Econometric Theory* **23**(5), 930–951.

Yan, Y., Gao, J. and Peng, B. (2020), Nonparametric time-varying vector moving average (∞) models. Working paper available at https://ssrn.com/abstract=3729872.

Zhang, T. and Wu, W. B. (2012), 'Inference of time-varying regression models', *Annals of Statistics* **40**(3), 1376–1402.

Zheng, Y. and Cheng, G. (2021), 'Finite-time analysis of vector autoregressive models under linear restrictions', *Biometrika* **108**(2), 469–489.

Supplementary material to "Parameter Stability Testing for Multivariate Dynamic Time-Varying Models"

In the supplementary file, Section S.1 shows that the nonparametric time-varying VARX process can be approximated by a time–varying vector moving average infinity (TV-VMA(∞)) process, which is the key to the proofs of the main results. In addition, we define notation and mathematical symblos for simplicity in this section. Section S.2 establishes the rates of (uniform) convergence for several estimators studied in this paper. Section S.3 provides the proofs of the main results. Some preliminary lemmas with their proofs are regulated to Section S.4 and Section S.5, respectively.

S.1 TV-VMA(∞) Representation & Notation

LEMMA S1. *Under Assumptions 1–5, there exists a TV-VMA(∞) process*

$$\widetilde{y}_t = \mu(\tau_t) + \sum_{j=0}^{\infty} D_j^{\epsilon}(\tau_t)\epsilon_{t-j} + \sum_{j=0}^{\infty} D_j^{v}(\tau_t)v_{t-j}$$

such that $\max_{t\geq 1}[E\{\|y_t - \widetilde{y}_t\|^{\delta}\}]^{1/\delta} = O(T^{-1})$*, where* $\mu(\tau) = \sum_{j=0}^{\infty}\sum_{l=0}^{q}\Psi_j(\tau)B_l(\tau)g(\tau)$*,* $D_{j,l}^{v}(\tau) = \sum_{k=0}^{j}\Psi_k(\tau)B_l(\tau)C_{j-k}(\tau)$*,* $D_j^{\epsilon}(\tau) = \Psi_j(\tau)\omega(\tau)$*,*
$D_j^{v}(\tau) = \sum_{b=\max(0,j-q)}^{j} D_{b,j-b}^{v}(\tau)$*,* $\Psi_j(\tau) = J\Phi^j(\tau)J^{\top}$*,* $J = \left[I_d, 0_{d\times d(p-1)}\right]$*, and*

$$\Phi(\tau) = \begin{pmatrix} A_1(\tau) & \cdots & A_{p-1}(\tau) & A_p(\tau) \\ I_d & \cdots & 0_d & 0_d \\ \vdots & \ddots & \vdots & \vdots \\ 0_d & \cdots & I_d & 0_d \end{pmatrix}.$$

Moreover, $\widetilde{y}_t$ *and* x_t *admit the following expression*

$$\begin{pmatrix} \widetilde{y}_t \\ x_t \end{pmatrix} = \begin{pmatrix} \mu(\tau_t) \\ g(\tau_t) \end{pmatrix} + \sum_{j=0}^{\infty} D_j(\tau_t) \begin{pmatrix} \epsilon_{t-j} \\ v_{t+1-j} \end{pmatrix},$$

where $D_j(\tau) = \begin{pmatrix} D_j^{\epsilon}(\tau) & D_{j-1}^{v}(\tau) \\ 0 & C_{j-1}(\tau) \end{pmatrix}$*, and* $D_j^{v}(\tau) = 0$ *and* $C_j(\tau) = 0$ *for* $j < 0$*. Here,* $D_j(\cdot)$*'s satisfy the same conditions as those in Assumption 3.*

Notation for Lemma S2: Define $\Sigma_z(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E(z_t z_t^{\top})K_h(\tau_t - \tau) = \begin{pmatrix} \Sigma_{z,11}(\tau) & \Sigma_{z,21}^{\top}(\tau) \\ \Sigma_{z,21}(\tau) & \Sigma_{z,22}(\tau) \end{pmatrix}$,

where

$$\Sigma_{z,11}(\tau) = \begin{pmatrix} \Sigma_{y,0}(\tau) & \cdots & \Sigma_{y,p}^{\top}(\tau) \\ \vdots & \ddots & \vdots \\ \Sigma_{y,p}(\tau) & \cdots & \Sigma_{y,0}(\tau) \end{pmatrix},$$

$$
\begin{aligned}
\Sigma_{y,k}(\tau) &= \mu(\tau)\mu(\tau)^\top + \sum_{j=0}^{\infty} D^{\epsilon}_{j-k}(\tau)D^{\epsilon}_{j}(\tau)^\top + \sum_{j=0}^{\infty} D^{v}_{j-k}(\tau)D^{v}_{j}(\tau)^\top \\
&\quad + \sum_{j=0}^{\infty} D^{\epsilon}_{j+1-k}(\tau)\rho_{\epsilon v}D^{v}_{j}(\tau)^\top + \left(\sum_{j=0}^{\infty} D^{\epsilon}_{j+1-k}(\tau)\rho_{\epsilon v}D^{v}_{j}(\tau)^\top\right)^\top, \\
\Sigma_{z,21}(\tau) &= \begin{pmatrix} \Sigma_{xy,-1}(\tau) & \cdots & \Sigma_{xy,p-1}(\tau) \\ \vdots & \ddots & \vdots \\ \Sigma_{xy,-q}(\tau) & \cdots & \Sigma_{xy,p-q}(\tau) \\ \Sigma_{xy,0}(\tau) & \cdots & \Sigma_{xy,p}(\tau) \end{pmatrix}, \\
\Sigma_{xy,k}(\tau) &= g(\tau)\mu(\tau)^\top + \sum_{j=0}^{\infty} C_{j+k}(\tau)\rho_{\epsilon v}^\top D^{\epsilon}_{j}(\tau)^\top + \sum_{j=0}^{\infty} C_{j+1+k}(\tau)D^{v}_{j}(\tau)^\top, \\
\Sigma_{z,22}(\tau) &= \begin{pmatrix} \Sigma_{x,0}(\tau) & \cdots & \Sigma_{x,q-1}^\top(\tau) & \Sigma_{x,1}(\tau) \\ \vdots & \ddots & \vdots & \vdots \\ \Sigma_{x,q-1}(\tau) & \cdots & \Sigma_{x,0}(\tau) & \Sigma_{x,q}(\tau) \\ \Sigma_{x,1}^\top(\tau) & \cdots & \Sigma_{x,q}^\top(\tau) & \Sigma_{x,0}(\tau) \end{pmatrix}, \\
\Sigma_{x,k}(\tau) &= g(\tau)g(\tau)^\top + \sum_{j=0}^{\infty} C_j(\tau)C_{j+k}^\top(\tau). \qquad \text{(S.1)}
\end{aligned}
$$

Here, the explicit structure of $\Sigma_z(\tau)$ follows directly from Lemma S1.

Additional notation: Let $\widetilde{c}_k = \int_{-1}^{1} u^k K(u)\mathrm{d}u$, $\Sigma_Z(\tau) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(Z_t Z_t^\top)K_h(\tau_t - \tau) = \Sigma_z(\tau) \otimes I_d$, $\Sigma = \int_0^1 (\Sigma_{X_C}(\tau) - \Sigma_{X_{C,\widetilde{C}}}(\tau)\Sigma^{-1}_{X_{\widetilde{C}}}(\tau)\Sigma^\top_{X_{C,\widetilde{C}}}(\tau))\mathrm{d}\tau$, where

$$
\begin{aligned}
\Sigma_{X_C}(\tau) &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(X_{C,t}X_{C,t}^\top)K_h(\tau_t - \tau) = C\Sigma_Z(\tau)C^\top, \\
\Sigma_{X_{C,\widetilde{C}}}(\tau) &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(X_{C,t}X_{\widetilde{C},t}^\top)K_h(\tau_t - \tau) = C\Sigma_Z(\tau)\widetilde{C}^\top, \\
\Sigma_{X_{\widetilde{C}}}(\tau) &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(X_{\widetilde{C},t}X_{\widetilde{C},t}^\top)K_h(\tau_t - \tau) = \widetilde{C}\Sigma_Z(\tau)\widetilde{C}^\top. \qquad \text{(S.2)}
\end{aligned}
$$

S.2 On the Rates of (Uniform) Convergence

LEMMA S2. *Suppose Assumptions 1–7 hold. As $T \to \infty$, then the following results hold.*

1. $\sup_{\tau\in[0,1]} \|\widehat{\beta}(\tau) - \beta(\tau)\| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{1/2}\right)$;

2. $\sup_{\tau\in[h,1-h]} \|\widehat{V}(\tau) - V(\tau)\| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{1/2}\right)$, *where* $V(\tau) = \Sigma_z^{-1}(\tau) \otimes \Omega(\tau)$.

It is noteworthy that the covariance matrix estimator $\widehat{V}(\tau)$ is constructed using a local constant kernel method rather than a local linear method, which is to avoid the non-positive definiteness issue associated with the local linear approach in finite sample study. Such a numerical problem has been well explained and investigated in the literature. We refer interested readers to Chen and Leng (2015) for details. The results of Lemma S2 provide uniform convergence rates for $\widehat{\beta}(\cdot)$ and $\widehat{V}(\cdot)$, which further facilitate the development for the test statistic.

LEMMA S3. *Let Assumptions 1–7 hold. As $T \to \infty$, $\sqrt{T}(\widehat{c} - c) = O_P(1)$.*

Lemma S3 indicates that the parametric component c can be estimated with a faster convergence rate. As a result, the asymptotic behaviour of $\widehat{Q}_{C,H}$ is solely determined by the nonparametric estimate $\widehat{\beta}(\tau)$.

S.3 Proofs of the Main Results

In what follows, M and $O(1)$ stand for constants, and may be different at each appearance.

Proof of Lemma S1. Define $J = \left[I_d, 0_{d\times d(p-1)}\right]$ and recall $\Psi_j(\tau) = J\Phi^j(\tau)J^\top$, where

$$\Phi(\tau) = \begin{pmatrix} A_1(\tau) & \cdots & A_{p-1}(\tau) & A_p(\tau) \\ I_d & \cdots & 0_d & 0_d \\ \vdots & \ddots & \vdots & \vdots \\ 0_d & \cdots & I_d & 0_d \end{pmatrix}.$$

To proceed, we write y_t as a TV-VMA(∞):

$$y_t = \sum_{j=0}^{\infty} \Psi_{j,t} \left(\sum_{l=0}^{q} B_l(\tau_{t-j}) x_{t-l-j} + \eta_{t-j} \right) = \mu_t + \sum_{j=0}^{\infty} D_{j,t}^{\epsilon} \epsilon_{t-j} + \sum_{l=0}^{q}\sum_{j=0}^{\infty} D_{j,l,t}^{v} v_{t-l-j},$$

where $\mu_t = \sum_{j=0}^{\infty}\sum_{l=0}^{q} \Psi_{j,t}B_l(\tau_{t-j})g(\tau_{t-l-j})$, $\Psi_{j,t} = J\prod_{m=0}^{j-1}\Phi(\tau_{t-m})J^\top$, $D_{j,t}^{\epsilon} = \Psi_{j,t}\omega(\tau_{t-j})$ and $D_{j,l,t}^{\boldsymbol{v}} = \sum_{k=0}^{j}\Psi_{k,t}B_l(\tau_{t-k})C_{j-k}(\tau_{t-l-k})$.

Let ρ denote the largest eigenvalue of $\Phi(\tau)$ uniformly over $\tau \in [0,1]$. Then we have $\rho < 1$ by Assumption 1. In addition, similar to the proof of Proposition 2.4 in Dahlhaus and Polonik (2009), we have $\max_{t\geq 1}\|\prod_{m=0}^{j-1}\Phi(\tau_{t-m})\| \leq M\rho^j$.

Next, we will show that y_t can be approximated by a TV-VMA(∞) process $\widetilde{y}_t$ satisfying $[E\{\|y_t - \widetilde{y}_t\|^{\delta}\}]^{1/\delta} = O(T^{-1})$, where $\widetilde{y}_t$ has been defined in the body of this lemma. By Minkowski inequality, we have

$$\begin{aligned}[E\{\|y_t - \widetilde{y}_t\|^{\delta}\}]^{1/\delta} &\leq M\left(\|\mu_t - \mu(\tau_t)\| + \sum_{j=0}^{\infty}\left\|D_{j,t}^{\epsilon} - D_j^{\epsilon}(\tau_t)\right\| + \sum_{l=0}^{q}\sum_{j=0}^{\infty}\left\|D_{j,l,t}^{\boldsymbol{v}} - D_{j,l}^{\boldsymbol{v}}(\tau_t)\right\| \right) \\ &= O(1)\cdot(I_{T,1} + I_{T,2} + I_{T,3}),\end{aligned}$$

where the definitions of $I_{T,1}$, $I_{T,2}$, and $I_{T,3}$ are obvious.

Consider $I_{T,1}$. Note that for any conformable matrices $\{A_i\}$ and $\{B_i\}$, since

$$\prod_{i=1}^{r} A_i - \prod_{i=1}^{r} B_i = \sum_{j=1}^{r}\left(\prod_{k=1}^{j-1} A_k\right)(A_j - B_j)\left(\prod_{k=j+1}^{r} B_k\right),$$

by Assumption 2, we obtain

$$\begin{aligned}
\|\Psi_{j,t}-\Psi_j(\tau_t)\| &= \left\| J\prod_{m=0}^{j-1}\Phi(\tau_{t-m})J^\top - J\Phi^j(\tau_t)J^\top \right\| \\
&\le M\sum_{i=1}^{j-1}\left\|\Phi^i(\tau_t)(\Phi(\tau_{t-i})-\Phi(\tau_t))\prod_{m=i+1}^{j-1}\Phi(\tau_{t-m})\right\| \le M\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1}.
\end{aligned}$$

Hence, by the Lipschitz continuity of $g(\tau)$, we have

$$\begin{aligned}
I_{T,1} &\le \sum_{j=0}^{\infty}\|\Psi_{j,t}-\Psi_j(\tau_t)\|\cdot\sum_{l=0}^{q}\|B_l(\tau_{t-j})g(\tau_{t-l-j})\| \\
&\quad+\sum_{j=0}^{\infty}\|\Psi_j(\tau_t)\|\cdot\sum_{l=0}^{q}\|B_l(\tau_{t-j})g(\tau_{t-l-j})-B_l(\tau_t)g(\tau_t)\| \\
&\le M\sum_{j=0}^{\infty}\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1}+M\sum_{j=0}^{\infty}\rho^j\frac{j}{T}=O(T^{-1}),
\end{aligned}$$

where we have used the facts that $\|\Psi_j(\tau)\|\le M\rho^j$ and

$$\begin{aligned}
&\|B_l(\tau_{t-j})g(\tau_{t-l-j})-B_l(\tau_t)g(\tau_t)\| \\
=&\|B_l(\tau_{t-j})g(\tau_{t-l-j})-B_l(\tau_t)g(\tau_{t-l-j})+B_l(\tau_t)g(\tau_{t-l-j})-B_l(\tau_t)g(\tau_t)\| \\
\le&\|B_l(\tau_{t-j})-B_l(\tau_t)\|\cdot\|g(\tau_{t-l-j})\|+\|B_l(\tau_t)\|\cdot\|g(\tau_{t-l-j})-g(\tau_t)\|\le M\frac{j}{T}.
\end{aligned}$$

Similarly, we have $I_{T,2}=O(T^{-1})$. For $I_{T,3}$,

$$\begin{aligned}
I_{T,3} &\le \sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\|\Psi_{k,t}B_l(\tau_{t-k})C_{j-k}(\tau_{t-l-k})-\Psi_k(\tau_t)B_l(\tau_t)C_{j-k}(\tau_t)\| \\
&=\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\Psi_{j,t}B_l(\tau_{t-j})C_k(\tau_{t-l-j})-\Psi_j(\tau_t)B_l(\tau_t)C_k(\tau_t)\| \\
&\le\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\Psi_{j,t}-\Psi_j(\tau_t)\|\cdot\|B_l(\tau_{t-j})\|\cdot\|C_k(\tau_{t-l-j})\| \\
&\quad+\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\Psi_j(\tau_t)\|\cdot\|B_l(\tau_{t-j})-B_l(\tau_t)\|\cdot\|C_k(\tau_{t-l-j})\| \\
&\quad+\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\Psi_j(\tau_t)\|\cdot\|B_l(\tau_t)\|\cdot\|C_k(\tau_{t-l-j})-C_k(\tau_t)\| \\
&\le M\sum_{j=0}^{\infty}\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1}\cdot\sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}\|C_k(\tau)\| \\
&\quad+\frac{M}{T}\left(\sum_{j=0}^{\infty}j\rho^j\right)\left(\sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}\|C_k(\tau)\|+\sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}\left\|C_k^{(1)}(\tau)\right\|\right)=O(T^{-1}).
\end{aligned}$$

Hence, we have proved $[E\|y_t-\widetilde{y}_t\|^\delta]^{1/\delta}=O(T^{-1})$. In addition, it is straightforward to verify that $\sup_{\tau\in[0,1]}\sum_{j=0}^{\infty}j\|D_j^{\epsilon,(k)}(\tau)\|<\infty$ and $\sup_{\tau\in[0,1]}\sum_{j=0}^{\infty}j\|D_{j,l}^{v,(k)}(\tau)\|<\infty$ for $k=0,1$. See the proofs of

Propositions 2.1 and 3.1 in Yan et al. (2020) for example. The proof is now complete. □

Proof of Lemma S2. (1). For notational simplicity, let $Z_{\tau,t}$ be the transpose of the tth row of Z_τ. Also, we define

$$
\begin{aligned}
S_{T,k}(\tau) &= \frac{1}{T}\sum_{t=1}^{T} Z_t Z_t^\top \left(\frac{\tau_t-\tau}{h}\right)^k K_h(\tau_t-\tau) \text{ for } 0\le k\le 3,\\
M(\tau_t) &= \beta(\tau_t)-\beta(\tau)-\beta^{(1)}(\tau)(\tau_t-\tau)-\frac{1}{2}\beta^{(2)}(\tau)(\tau_t-\tau)^2,\\
S_T(\tau) &= \begin{pmatrix} S_{T,0}(\tau) & S_{T,1}(\tau)\\ S_{T,1}(\tau) & S_{T,2}(\tau)\end{pmatrix}.
\end{aligned}
$$

Since $y_t = Z_t^\top\left(\beta(\tau)+\beta^{(1)}(\tau)(\tau_t-\tau)+\frac{1}{2}\beta^{(2)}(\tau)(\tau_t-\tau)^2+M(\tau_t)\right)+\eta_t$, we can write

$$
\begin{aligned}
\widehat{\beta}(\tau)-\beta(\tau) &= [I_l,0_l]\left(\frac{1}{T}\sum_{t=1}^{T} Z_{\tau,t}Z_{,\tau,t}^\top K_h(\tau_t-\tau)\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_{\tau,t}y_t K_h(\tau_t-\tau)-\beta(\tau)\\
&= [I_l,0_l]S_T^{-1}(\tau)\begin{bmatrix} S_{T,2}(\tau)\\ S_{T,3}(\tau)\end{bmatrix}\frac{1}{2}h^2\beta^{(2)}(\tau)+[I_l,0_l]S_T^{-1}(\tau)\cdot\frac{1}{T}\sum_{t=1}^{T} Z_{\tau,t}Z_t^\top M(\tau_t)K_h(\tau_t-\tau)\\
&\quad +[I_l,0_l]S_T^{-1}(\tau)\cdot\frac{1}{T}\sum_{t=1}^{T} Z_{\tau,t}\eta_t K_h(\tau_t-\tau).
\end{aligned}
$$

The uniform convergence rate for $\widehat{\beta}(\tau)$ follows directly from Lemmas S4 and S6.3.

(2). By Lemma S4, we have

$$
\sup_{\tau\in[h,1-h]}\left\|\widehat{\Sigma}_z(\tau)-\Sigma_z(\tau)\right\| = O_P\left(h^2+\left(\frac{\log T}{Th}\right)^{1/2}\right).
$$

Then we just need to focus on the rate associated with $\widehat{\Omega}(\tau)$. For notational simplicity, we ignore the $\frac{1}{T}\sum_{t=1}^{T}K_h(\tau_t-\tau)$ as $\frac{1}{T}\sum_{t=1}^{T}K_h(\tau_t-\tau)=1+O((Th)^{-1})$ uniformly over $\tau\in[h,1-h]$. Write

$$
\begin{aligned}
\widehat{\Omega}(\tau) &= \frac{1}{Th}\sum_{t=1}^{T}\widehat{\eta}_t\widehat{\eta}_t^\top K\left(\frac{\tau_t-\tau}{h}\right) = \frac{1}{Th}\sum_{t=1}^{T}\left(\eta_t+\widehat{\eta}_t-\eta_t\right)\left(\eta_t+\widehat{\eta}_t-\eta_t\right)^\top K\left(\frac{\tau_t-\tau}{h}\right)\\
&= \frac{1}{Th}\sum_{t=1}^{T}\eta_t\eta_t^\top K\left(\frac{\tau_t-\tau}{h}\right)+\frac{1}{Th}\sum_{t=1}^{T}(\widehat{\eta}_t-\eta_t)(\widehat{\eta}_t-\eta_t)^\top K\left(\frac{\tau_t-\tau}{h}\right)\\
&\quad +\frac{1}{Th}\sum_{t=1}^{T}\eta_t(\widehat{\eta}_t-\eta_t)^\top K\left(\frac{\tau_t-\tau}{h}\right)+\frac{1}{Th}\sum_{t=1}^{T}(\widehat{\eta}_t-\eta_t)\eta_t^\top K\left(\frac{\tau_t-\tau}{h}\right)\\
&= I_{T,1}+I_{T,2}+I_{T,3}+I_{T,4}.
\end{aligned}
$$

Consider $I_{T,1}$. By Lemma S6.4, we have

$$
\sup_{\tau\in[0,1]}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\eta_t\eta_t^\top-E\left(\eta_t\eta_t^\top\right)\right]K_h(\tau_t-\tau)\right\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right).
$$

Next, consider $I_{T,2}$. By Lemma S4, we have

$$\sup_{\tau\in[0,1]}\frac{1}{T}\sum_{t=1}^{T}\|Z_t\|^2 K_h(\tau_t-\tau)\le \sup_{\tau\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\operatorname{tr}\left[Z_t^\top Z_t-E(Z_t^\top Z_t)\right]K_h(\tau_t-\tau)\right|$$
$$+\sup_{\tau\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{T}\operatorname{tr}\left[E(Z_t^\top Z_t)\right]K_h(\tau_t-\tau)\right|=o_P(1)+O(1)=O_P(1).$$

Hence, by the first result of this lemma

$$\sup_{\tau\in[0,1]}\|I_{T,2}\|\le\sup_{\tau\in[0,1]}\|\widehat{\beta}(\tau)-\beta(\tau)\|^2\cdot\sup_{\tau\in[0,1]}\frac{1}{T}\sum_{t=1}^{T}\|Z_t\|^2 K_h(\tau_t-\tau)=o_P\left(h^2+\sqrt{\frac{\log T}{Th}}\right).$$

Similarly, for $I_{T,3}$ and $I_{T,4}$, we have

$$\sup_{\tau\in[0,1]}\|I_{T,3}\|\le\sup_{\tau\in[0,1]}\left\|\widehat{\beta}(\tau)-\beta(\tau)\right\|\cdot\sup_{\tau\in[0,1]}\frac{1}{T}\sum_{t=1}^{T}\|Z_t\eta_t\|K_h(\tau_t-\tau)$$
$$\le\sup_{\tau\in[0,1]}\left\|\widehat{\beta}(\tau)-\beta(\tau)\right\|\cdot\left\{\sup_{\tau\in[0,1]}\frac{1}{T}\sum_{t=1}^{T}\|Z_t\|^2K_h(\tau_t-\tau)\right\}^{1/2}$$
$$\cdot\left\{\sup_{\tau\in[0,1]}\frac{1}{T}\sum_{t=1}^{T}\|\eta_t\|^2K_h(\tau_t-\tau)\right\}^{1/2}=O_P\left(h^2+\sqrt{\frac{\log T}{Th}}\right).$$

The proof of the second result is now completed. □

Proof of of Lemma S3. By Lemma S7,

$$\sqrt{T}\left(\widehat{c}-c\right)$$
$$=\left(X_C^\top(I_{dT}-S)^\top(I_{dT}-S)X_C\right)^{-1}X_C^\top(I_{dT}-S)^\top(I_{dT}-S)\left(\begin{bmatrix}X_{\widetilde{C},1}^\top\theta(\tau_1)\\ \vdots\\ X_{\widetilde{C},T}^\top\theta(\tau_T)\end{bmatrix}+\eta\right)$$
$$=\Sigma^{-1}\frac{1}{\sqrt{T}}X_C^\top(I_{dT}-S)^\top(I_{dT}-S)\eta+o_P(1).$$

Hence, it suffices to show that $\frac{1}{\sqrt{T}}X_C^\top(I_{dT}-S)^\top(I_{dT}-S)\eta=O_P(1)$.

By using the same arguments as the proof of Lemma S7,

$$\frac{1}{\sqrt{T}}X_C^\top(I_{dT}-S)^\top(I_{dT}-S)\eta$$
$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(X_{C,t}-\Sigma_{X_{C,\widetilde{C}}}(\tau_t)\Sigma_{X_{\widetilde{C}}}^{-1}(\tau_t)X_{\widetilde{C},t}\right)\eta_t+o_P(1).$$

Since $\left\{\left(X_{C,t}-\Sigma_{X_{C,\widetilde{C}}}(\tau_t)\Sigma_{X_{\widetilde{C}}}^{-1}(\tau_t)X_{\widetilde{C},t}\right)\eta_t\right\}$ is a martingale difference sequence, the result follows directly. □

Proof of Theorem 1. First, we denote a few notations to facilitate the development. Let $A_Q=\widetilde{v}_0\cdot\operatorname{tr}\left\{\int_0^1\Sigma_Q(\tau)\mathrm{d}\tau\right\}$ and $B_Q=4C_B\cdot\operatorname{tr}\left\{\int_0^1\Sigma_Q(\tau)^2\mathrm{d}\tau\right\}$, where $\Sigma_Q(\tau)=H(\tau)^{1/2}CV(\tau)C^\top H(\tau)^{1/2}$. Let

$\rho_T = h^2 + \sqrt{\frac{\log T}{Th}}$. By Lemma S2, we have

$$\sup_{\tau\in[0,1]} \left\|\widehat{\beta}(\tau) - \beta(\tau)\right\| = O_P\left(\rho_T\right). \tag{S.3}$$

Then we can conclude that as $Th^{11/2} = o(1)$,

$$\int_{\mathcal{B}_T} \left[C\widehat{\beta}(\tau) - c\right]^\top H(\tau) \left[C\widehat{\beta}(\tau) - c\right] \mathrm{d}\tau = O_P\left(\log T/T + h^5\right) = o_P\left(T^{-1}h^{-1/2}\right),$$

where $\mathcal{B}_T = [0, h] \cup [1-h, 1]$.

In addition, by Lemma S4 and Lemma S6.3, we have

$$\begin{aligned}
&\sup_{[h,1-h]} \|S_T(\tau) - \Sigma_Z(\tau) \otimes \Lambda_1\| = +O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right), \\
&\sup_{[0,1]} \|R_T(\tau)\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right),
\end{aligned} \tag{S.4}$$

where $\Lambda_1 = \mathrm{diag}(\widetilde{c}_0, \widetilde{c}_2)$ and $R_T(\tau) = \frac{1}{T}\sum_{t=1}^T Z_t\eta_t K_h(\tau_t - \tau)$. Hence, under the null, we have

$$\sup_{[h,1-h]} \left\|C\widehat{\beta}(\tau) - c - C\Sigma_Z^{-1}(\tau)R_T(\tau)\right\| = O_P(\rho_T^2). \tag{S.5}$$

By (S.4) and $Th^{11/2} \to 0$, we have

$$\begin{aligned}
&\int_0^1 \left[C\widehat{\beta}(\tau) - c\right]^\top H(\tau) \left[C\widehat{\beta}(\tau) - c\right] \mathrm{d}\tau \\
&= \int_h^{1-h} R_T^\top(\tau)H_0(\tau)R_T(\tau)\mathrm{d}\tau + O_P\left(\log T/T + h^5\right) + O_P\left(\rho_T^2\sqrt{\frac{\log T}{Th}}\right) \\
&= \int_0^1 R_T^\top(\tau)H_0(\tau)R_T(\tau)\mathrm{d}\tau + o_P(T^{-1}h^{-1/2}),
\end{aligned}$$

where $H_0(\tau) = \Sigma_Z^{-1}(\tau)C^\top H(\tau)C\Sigma_Z^{-1}(\tau)$.

Consider $\int_0^1 R_T^\top(\tau)H_0(\tau)R_T(\tau)\mathrm{d}\tau$, and write

$$\begin{aligned}
&\int_0^1 R_T^\top(\tau)H_0(\tau)R_T(\tau)\mathrm{d}\tau = \frac{1}{T^2h^2}\sum_{t=1}^T \eta_t^\top Z_t^\top \left\{\int_0^1 H_0(\tau)K^2\left(\frac{\tau-\tau_t}{h}\right)\mathrm{d}\tau\right\} Z_t\eta_t \\
&+\frac{1}{T^2h^2}\sum_{t=1}^T\sum_{s=1,\neq t}^T \eta_t^\top Z_t^\top \left\{\int_0^1 H_0(\tau)K\left(\frac{\tau-\tau_t}{h}\right)K\left(\frac{\tau-\tau_s}{h}\right)\mathrm{d}\tau\right\} Z_s\eta_s \\
&= I_{T,1} + I_{T,2},
\end{aligned}$$

where the definitions of $I_{T,1}$ and $I_{T,2}$ should be obvious.

For $I_{T,1}$, simple algebra shows that

$$\eta_t^\top Z_t^\top \Sigma_Z^{-1}(\tau)C^\top H(\tau)C\Sigma_Z^{-1}(\tau)Z_t\eta_t = \mathrm{tr}\left\{\left[\left(\Sigma_z^{-1}(\tau)z_tz_t^\top\Sigma_z^{-1}(\tau)\right)\otimes\eta_t\eta_t^\top\right]C^\top H(\tau)C\right\}.$$

Then we can have

$$
\begin{aligned}
I_{T,1} &= \widetilde{v}_0 \frac{1}{T^2 h} \sum_{t=1}^{T} \eta_t^\top Z_t^\top \left[H_0(\tau_t) + O(h) \right] Z_t \eta_t \\
&= \widetilde{v}_0 \frac{1}{T^2 h} \sum_{t=1}^{T} \operatorname{tr} \left\{ \left[\left(\Sigma_z^{-1}(\tau_t) z_t z_t^\top \Sigma_z^{-1}(\tau_t) \right) \otimes \eta_t \eta_t^\top \right] \cdot C^\top H(\tau_t) C \right\} + O_P(T^{-1}) \\
&= \widetilde{v}_0 \frac{1}{T^2 h} \sum_{t=1}^{T} \operatorname{tr} \left\{ \left[\Sigma_z^{-1}(\tau_t) \otimes \Omega(\tau_t) \right] \cdot C^\top H(\tau_t) C \right\} + O_P(T^{-1} + T^{-3/2} h^{-1}) \\
&\to_P (Th)^{-1} A_Q.
\end{aligned}
$$

Next, consider $I_{T,2}$. Let $w_{s,t} = \frac{1}{T\sqrt{h}} \int_{-1}^{1} K(u) K\left(u + \frac{t-s}{Th}\right) \mathrm{d}u$. Since

$$
\int_0^1 H_0(\tau) K\left(\frac{\tau - \tau_t}{h}\right) K\left(\frac{\tau - \tau_s}{h}\right) \mathrm{d}\tau = h \int_{-1}^{1} H_0(\tau_t + uh) K(u) K\left(u + \frac{t-s}{Th}\right) \mathrm{d}u,
$$

we have $T\sqrt{h} I_{T,2} = 2\sum_{t=2}^{T} \sum_{s=1}^{t-1} \eta_t^\top Z_t^\top H_0(\tau_t) Z_s \eta_s w_{s,t}(1 + o(1)) = 2\widetilde{U} + o_P(1)$, where the definition of $\widetilde{U}$ is obvious. By Lemma S12, we have

$$
\widetilde{U} \to_D N\left(0, \sigma_{\widetilde{U}}^2\right),
$$

where $\sigma_{\widetilde{U}}^2 = \lim_{T\to\infty} \sum_{t=2}^{T} \operatorname{tr} \left\{ E\left(H_0^\top(\tau_t) Z_t \eta_t \eta_t^\top Z_t^\top H_0(\tau_t) \right) E\left(\sum_{s=1}^{t-1} Z_s \eta_s \eta_s^\top Z_s^\top \right) w_{s,t}^2 \right\}$.

Next, we show that $\sigma_{\widetilde{U}}^2 = C_B \operatorname{tr} \left\{ \int_0^1 \Sigma_Q(\tau)^2 \mathrm{d}\tau \right\}$. Let $V_1(\tau) = H_0(\tau) V_2(\tau) H_0(\tau)$ and $V_2(\tau) = \Sigma_z(\tau) \otimes \Omega(\tau)$. Write

$$
\begin{aligned}
&\sum_{t=2}^{T} E\left(H_0^\top(\tau_t) Z_t \eta_t \eta_t^\top Z_t^\top H_0(\tau_t) \right) E\left(\sum_{s=1}^{t-1} Z_s \eta_s \eta_s^\top Z_s^\top \right) w_{s,t}^2 \\
&= \sum_{t=2}^{T} \sum_{s=1}^{t-1} V_1(\tau_t) V_2(\tau_s) w_{s,t}^2 = \frac{1}{T^2 h} \sum_{t=2}^{T} \sum_{s=1}^{t-1} V_1(\tau_t) V_2(\tau_s) \left[\int_{-1}^{1} K(u) K\left(u + \frac{t-s}{Th}\right) \mathrm{d}u \right]^2 \\
&= \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} V_1(\tau_s + j/T) V_2(\tau_s) \left[\int_{-1}^{1} K(u) K\left(u + \frac{t-s}{Th}\right) \mathrm{d}u \right]^2 \\
&= \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} V_1(\tau_s + j/T) V_2(\tau_s) \left[\int_{-1}^{1} K(u) K\left(u + \frac{j}{Th}\right) \mathrm{d}u \right]^2 \\
&= \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} V_1(\tau_s) V_2(\tau_s) \left[\int_{-1}^{1} K(u) K\left(u + \frac{j}{Th}\right) \mathrm{d}u \right]^2 \\
&\quad + \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} O(j/T) V_2(\tau_s) \left[\int_{-1}^{1} K(u) K\left(u + \frac{j}{Th}\right) \mathrm{d}u \right]^2 = I_{T,3} + I_{T,4},
\end{aligned}
$$

where the definitions of $I_{T,3}$ and $I_{T,4}$ are obvious.

It is easy to verify $\operatorname{tr}\{I_{T,3}\} \to C_B \operatorname{tr} \left\{ \int_0^1 \Sigma_Q(\tau)^2 \mathrm{d}\tau \right\}$. For $I_{T,4}$, we have

$$
\|I_{T,4}\| \le M \frac{1}{Th} \sum_{j=1}^{T} j/T \left[\int_{-1}^{1} K(u) K\left(u + \frac{j}{Th}\right) \mathrm{d}u \right]^2
$$

$$\simeq M\int_0^2 vh\left[\int_{-1}^1 K(u)K(u+v)\mathrm{d}u\right]^2\mathrm{d}v=o(1).$$

Combining the above results, we have proved

$$T\sqrt{h}\left[\int_0^1 R_T^\top(\tau)H_0(\tau)R_T(\tau)\mathrm{d}\tau-(Th)^{-1}A_Q\right]\to_D N(0,B_Q).$$

Note that

$$\begin{aligned}&\int_0^1\left[C\widehat{\beta}(\tau)-\widehat{c}\right]^\top H(\tau)\left[C\widehat{\beta}(\tau)-\widehat{c}\right]\mathrm{d}\tau-\int_0^1\left[C\widehat{\beta}(\tau)-c\right]^\top H(\tau)\left[C\widehat{\beta}(\tau)-c\right]\mathrm{d}\tau\\&=\int_0^1(\widehat{c}-c)^\top H(\tau)(\widehat{c}-c)\mathrm{d}\tau-2\int_0^1(\widehat{c}-c)^\top H(\tau)\left(C\widehat{\beta}(\tau)-c\right)\mathrm{d}\tau=I_{T,5}-2I_{T,6},\end{aligned}$$

where the definitions of $I_{T,5}$ and $I_{T,6}$ are obvious.

Since $\widehat{c}=c+O_P\left(T^{-1/2}\right)$, we have $I_{T,5}=O_P(T^{-1})$. For $I_{T,6}$, by (S.3) and (S.5), we have

$$\begin{aligned}I_{T,6}&=(\widehat{c}-c)^\top\int_h^{1-h}H(\tau)C\Sigma_Z^{-1}(\tau)R_T(\tau)\mathrm{d}\tau+o_P\left(T^{-1}h^{-1/2}\right)\\&=O_P(T^{-1})+o_P\left(T^{-1}h^{-1/2}\right)=o_P\left(T^{-1}h^{-1/2}\right)\end{aligned}$$

provided that

$$\begin{aligned}&\int_h^{1-h}H(\tau)C\Sigma_Z^{-1}(\tau)R_T(\tau)\mathrm{d}\tau\\&=\frac{1}{T}\sum_{t=1}^T\int_{-1}^1 H(\tau_t+uh)C\Sigma_Z^{-1}(\tau_t+uh)K(u)\mathrm{d}uZ_t\eta_t=O_P(T^{-1/2}).\end{aligned}$$

Now, we conclude that $T\sqrt{h}\left[\widehat{Q}_{C,H}-(Th)^{-1}A_Q\right]\to_D N(0,B_Q)$.

Now, write

$$\begin{aligned}&\int_0^1\left[C\widehat{\beta}(\tau)-\widehat{c}\right]^\top\widehat{H}(\tau)\left[C\widehat{\beta}(\tau)-\widehat{c}\right]\mathrm{d}\tau-\int_0^1\left[C\widehat{\beta}(\tau)-c\right]^\top H(\tau)\left[C\widehat{\beta}(\tau)-c\right]\mathrm{d}\tau\\&=\int_0^1\left[C\widehat{\beta}(\tau)-\widehat{c}\right]^\top\widehat{H}(\tau)\left[C\widehat{\beta}(\tau)-\widehat{c}\right]\mathrm{d}\tau-\int_0^1\left[C\widehat{\beta}(\tau)-\widehat{c}\right]^\top H(\tau)\left[C\widehat{\beta}(\tau)-\widehat{c}\right]\mathrm{d}\tau\\&\quad+\int_0^1\left[C\widehat{\beta}(\tau)-\widehat{c}\right]^\top H(\tau)\left[C\widehat{\beta}(\tau)-\widehat{c}\right]\mathrm{d}\tau-\int_0^1\left[C\widehat{\beta}(\tau)-c\right]^\top H(\tau)\left[C\widehat{\beta}(\tau)-c\right]\mathrm{d}\tau.\end{aligned}$$

Then we just need to focus on

$$I_{T,7}=\int_0^1\left[C\widehat{\beta}(\tau)-\widehat{c}\right]^\top\left(\widehat{H}(\tau)-H(\tau)\right)\left[C\widehat{\beta}(\tau)-\widehat{c}\right]\mathrm{d}\tau.$$

Hence, it suffices to show $T\sqrt{h}I_{T,7}=o_P(1)$. Using Lemma S2, it is easy to know that

$$\begin{aligned}|I_{T,7}|&\leq\sup_{\tau\in[0,1]}\left\|\widehat{H}(\tau)-H(\tau)\right\|\times\widehat{Q}_{C,I_s}\\&=O_P\left(h+\sqrt{\frac{\log T}{Th}}\right)O_P\left((Th)^{-1}+1/(T\sqrt{h})\right)=o_P(1/(T\sqrt{h})).\end{aligned}$$

The proof is now completed. □

Proof of Theorem 2. Under the local alternative (2.11), we have $C\beta(\tau) = c + d_T f(\tau)$ and thus

$$
\begin{aligned}
&\widehat{Q}_{C,H} - \int_0^1 R_T^\top(\tau)H_0(\tau)R_T(\tau)\mathrm{d}\tau \\
&= d_T^2 \int_0^1 f(\tau)^\top H(\tau) f(\tau)\mathrm{d}\tau + 2d_T \int_0^1 f(\tau)^\top H(\tau)\left(C\widehat{\beta}(\tau) - C\beta(\tau)\right)\mathrm{d}\tau \\
&\quad + \left[\int_0^1 \left(C\widehat{\beta}(\tau) - C\beta(\tau)\right)^\top H(\tau)\left(C\widehat{\beta}(\tau) - C\beta(\tau)\right)\mathrm{d}\tau - \int_0^1 R_T^\top(\tau)H_0(\tau)R_T(\tau)\mathrm{d}\tau\right] \\
&= d_T^2 \int_0^1 f(\tau)^\top H(\tau) f(\tau)\mathrm{d}\tau + I_{T,1} + I_{T,2}.
\end{aligned}
$$

Since $C\widehat{\beta}(\tau) - C\beta(\tau) = O_P\left(d_T\rho_T + \sqrt{\frac{\log T}{Th}}\rho_T\right) + C\Sigma_Z^{-1}(\tau)R_T(\tau)$ uniformly over $\tau \in [h, 1-h]$ and

$$
\int_0^1 f(\tau)^\top H(\tau) C\Sigma_Z^{-1}(\tau)R_T(\tau)\mathrm{d}\tau = O_P(T^{-1/2}),
$$

we have $I_{T,1} = O_P\left(d_T(d_T\rho_T + \sqrt{\frac{\log T}{Th}}\rho_T + T^{-1/2})\right) = o_P(T^{-1}h^{-1/2})$.

For $I_{T,2}$, since $\sup_{\tau\in[0,1]}\|R_T(\tau)\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right)$, we have

$$
I_{T,2} = O_P\left(d_T^2\rho_T^2 + d_T\rho_T\sqrt{\frac{\log T}{Th}}\right) = o_P(T^{-1}h^{-1/2}).
$$

As $T\sqrt{h}\left(\int_0^1 R_T^\top(\tau)H_0(\tau)R_T(\tau) - (Th)^{-1}A_Q\right) \to N(0, B_Q)$, we have

$$
T\sqrt{h}\left(\widehat{Q}_{C,H} - (Th)^{-1}A_Q\right) \to N(\delta_1, B_Q).
$$

In addition, similar to the proofs of Theorem 1, we have $T\sqrt{h}\left(\widehat{Q}_{C,H} - \widehat{Q}_{C,\widehat{H}}\right) = o_P(1)$. The proof is now complete. □

S.4 Preliminary Lemmas

For notational simplicity, we first define a few notations which will be repeatedly used below. Let $a_t = \sum_{s=1}^{t-1} w_{s,t}^2$, $b_s = \sum_{t=s+1}^{T} w_{s,t}^2$ and $\sigma_T^2 = \sum_{t=2}^{T} a_t$, where $w_{s,t} = \frac{1}{T\sqrt{h}}\int_{-1}^{1} K(u)\, K\left(u + \frac{t-s}{Th}\right)\mathrm{d}u$. For a random vector z, we write $z \in \mathcal{L}^q$, $q > 0$, if $|||z|||_q = \left[E\left(\|z\|^q\right)\right]^{1/q} < \infty$, and denote $|||\cdot||| = |||\cdot|||_2$ for short. Let $z_{t-1}^* = \sum_{s=1}^{t-1} w_{s,t} y_s^*$, where $y_t^* \in \mathcal{L}^\delta$, $t = 1, \ldots, T$ are martingale differences subject to the filtration $\mathcal{F}_t$ and $\delta > 4$. Let $Q_T = \sum_{t=2}^{T} y_t^{*,\top} H_t z_{t-1}^*$, where H_t is a deterministic weighting matrix.

LEMMA S4. Let Assumptions 3, 5 and 7 hold. In addition, let $\{W_{T,t}(\cdot)\}_{t=1}^{T}$ be a sequence of $q \times m$ matrices of deterministic functions, in which q is fixed, each functional component is Lipschitz continuous and defined on a compact set $[a, b]$. Moreover, suppose that

1. $\sup_{\tau\in[a,b]} \sum_{t=1}^{T} \|W_{T,t}(\tau)\| = O(1)$;
2. $\sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} \|W_{T,t+1}(\tau) - W_{T,t}(\tau)\| = O(d_T)$ with $d_T = \sup_{\tau\in[a,b],t\geq 1} \|W_{T,t}(\tau)\|$.

Then as $T \to \infty$,

1. $\sup_{\tau\in[a,b]} \| \sum_{t=1}^{T} W_{T,t}(\tau)\,(x_t - E(x_t))\,\| = O_P(\sqrt{d_T \log T})$ provided $T^{\frac{2}{\delta}} d_T \log T \to 0$;

2. $\sup_{\tau\in[a,b]} \| \sum_{t=1}^{T} W_{T,t}(\tau)\,\big(x_t x_{t+p}^{\top} - E(x_t x_{t+p}^{\top})\big)\,\| = O_P(\sqrt{d_T \log T})$ for any fixed integer $p \geq 0$ provided $T^{\frac{4}{\delta}} d_T \log T \to 0$, where δ is the same as in Assumption 1.5.

Lemma S4 is the Theorem 2.1 in Yan et al. (2020).

LEMMA S5. Let $\{Z_t, \mathcal{F}_t\}$ be a martingale difference sequence. Suppose that $|Z_t| \leq M$ for a constant M, $t = 1, \ldots, T$. Let $V_T = \sum_{t=1}^{T} \mathrm{Var}\,(Z_t \mid \mathcal{F}_{t-1}) \leq V$ for some $V > 0$. Then for any given $\nu > 0$,

$$\mathrm{pr}\left(\left|\sum_{t=1}^{T} Z_t\right| > \nu\right) \leq \exp\left\{-\frac{\nu^2}{2(V + M\nu)}\right\}.$$

Lemma S5 is Proposition 2.1 of Freedman (1975).

LEMMA S6. Let Assumptions 1–7 hold. As $T \to \infty$,

1. $\sup_{\tau\in[h,1-h]} \left\| \frac{1}{T} X_{\widetilde{C},\tau}^{\top} K_\tau X_{\widetilde{C},\tau} - \Sigma_{X_{\widetilde{C}}}(\tau) \otimes \Lambda_1 \right\| = O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right)$ with $\Lambda_1 = \mathrm{diag}(\tilde{c}_0, \tilde{c}_2)$;

2. $\sup_{\tau\in[h,1-h]} \left\| \frac{1}{T} X_{\widetilde{C},\tau}^{\top} K_\tau X_{C,\tau} - \Sigma_{X_{C,\widetilde{C}}}^{\top}(\tau) \otimes \Lambda_2 \right\| = O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right)$ with $\Lambda_2 = (\tilde{c}_0, 0)^{\top}$;

3. $\sup_{\tau\in[0,1]} \left\| \frac{1}{T} Z_\tau K_\tau \eta \right\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right)$, where $\eta = (\eta_1, \ldots, \eta_T)^{\top}$;

4. $\sup_{\tau\in[0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T} (\eta_t \eta_t^{\top} - E(\eta_t \eta_t^{\top})) K_h(\tau_t - \tau) \right\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right)$.

LEMMA S7. Let Assumptions 1–7 hold. As $T \to \infty$,

1. $\frac{1}{T} X_C^{\top} (I_{dT} - S)^{\top} (I_{dT} - S) X_C = \Sigma + o_P(1)$,
   where $\Sigma = \int_0^1 \Sigma_{X_C}(\tau) \mathrm{d}\tau - \int_0^1 \Sigma_{X_{C,\widetilde{C}}}(\tau) \Sigma_{X_{\widetilde{C}}}^{-1}(\tau) \Sigma_{X_{C,\widetilde{C}}}^{\top}(\tau) \mathrm{d}\tau$;

2. $X_C^{\top} (I_{dT} - S)^{\top} (I_{dT} - S) \widetilde{X} = o_P(\sqrt{T})$,
   where $\widetilde{X} = ([X_{\widetilde{C},1}^{\top} \theta(\tau_1)]^{\top}, \ldots, [X_{\widetilde{C},T}^{\top} \theta(\tau_T)]^{\top})^{\top}$.

LEMMA S8. Let Assumption 6 hold. Then the following results hold.

1. $\sigma_T^2 \to \int_0^2 \left[\int_{-1}^{1-v} K\,(u)\, K\,(u+v)\, \mathrm{d}u\right]^2 \mathrm{d}v$;

2. $\max_{2\leq t\leq T} a_t = O(1/T)$, where $a_t = \sum_{s=1}^{t-1} w_{s,t}^2$;

3. for $\forall J \in \mathbb{N}$, $\sum_{s=1}^{T-J} w_{s,s+J}^2 = O(1/(Th))$;

4. $T \sum_{s=1}^{T-1} b_s^2 = O(1)$, where $b_s = \sum_{t=s+1}^{T} w_{s,t}^2$;

5. $\sum_{k=1}^{T-1} \sum_{t=1}^{k-1} \left(\sum_{j=k+1}^{T} w_{k,j} w_{t,j}\right)^2 = O(1/T)$.

LEMMA S9. Let $\xi_1,...,\xi_T$ be a d-dimensional martingale difference for which $\xi_t \in \mathcal{L}^p$, $p > 1$. Let $p^* = \min(2, p)$. Then

$$\left\|\left\|\sum_{t=1}^{T} \xi_t\right\|\right\|_p^{p^*} \leq M \sum_{t=1}^{T} \|\|\xi_t\|\|_p^{p^*}.$$

LEMMA S10. Let Assumptions 5–6 hold. Assume $w_t^* \in \mathcal{L}^{\delta/2}$ for some $\delta > 4$ is $\mathcal{F}_t$-measurable. As $T \to \infty$,

$$E\left|\sum_{t=2}^{T} \operatorname{tr}\left[\left(w_t^* - E(w_t^*)\right) z_{t-1}^* z_{t-1}^{*,\top}\right]\right| \to 0.$$

LEMMA S11. Let Assumptions 1–6 hold and $y_t^* = Z_t \eta_t$. As $T \to \infty$,

$$\sum_{t=2}^{T} \operatorname{tr}\left[H_t\left(z_{t-1}^* z_{t-1}^{*,\top} - E(z_{t-1}^* z_{t-1}^{*,\top})\right)\right] \to_P 0,$$

where H_t is a deterministic weighting function satisfying $\|H_t\| < \infty$.

LEMMA S12. Let Assumptions 1–6 holds and $y_t^* = Z_t \eta_t$. As $T \to \infty$,

$$Q_T \to_D N\left(0, \sigma_Q^2\right).$$

where $\sigma_Q^2 = \lim_{T\to\infty} \sum_{t=2}^{T} \operatorname{tr}\left[E(H_t^\top y_t^* y_t^{*,\top} H_t) E(z_{t-1}^* z_{t-1}^{*,\top})\right]$.

S.5 Proofs of the Preliminary Lemmas

Proof of Lemma S6. (1)-(2). To prove parts (1) and (2), it suffices to show that

$$\sup_{\tau\in[0,1]} \left\|\frac{1}{T}\sum_{t=1}^{T}\left(Z_t Z_t^\top - E(Z_t Z_t^\top)\right)\left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau)\right\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right)$$

for $k = 0, 1, 2$. Since Z_t can be approximated by a TV-VMA(∞) process by Lemma S1, then the uniform convergence results follows directly from Lemma S4.

(3). Part (3) follows directly from Lemma B.8 in Yan et al. (2020).

(4). We use a number of subintervals $\{S_l\}$ to cover the interval $[0, 1]$, which are centered at s_l with the length $\delta_T = o(h^2)$. Denote the number of these intervals by N_T, then $N_T = O(\delta_T^{-1})$. Hence,

$$\begin{aligned}
&\sup_{\tau\in[0,1]} \left\|\frac{1}{T}\sum_{t=1}^{T}\left[\eta_t\eta_t^\top - E(\eta_t\eta_t^\top)\right] K_h(\tau_t - \tau)\right\| \\
\leq &\max_{1\leq l\leq N_T} \left\|\frac{1}{T}\sum_{t=1}^{T}\left[\eta_t\eta_t^\top - E(\eta_t\eta_t^\top)\right] K_h(\tau_t - s_l)\right\| \\
&+ \max_{1\leq l\leq N_T} \sup_{\tau\in S_l} \left\|\frac{1}{T}\sum_{t=1}^{T}\left[\eta_t\eta_t^\top - E(\eta_t\eta_t^\top)\right] (K_h(\tau_t - s_l) - K_h(\tau_t - \tau))\right\| I_{T,1} + I_{T,2},
\end{aligned}$$

where the definitions of $I_{T,1}$ and $I_{T,2}$ are obvious.

Let $\delta_T = O(\gamma_T h^2)$ with $\gamma_T = \sqrt{\frac{\log T}{Th}}$. Then we have

$$E|I_{T,2}| \leq M \frac{\delta_T}{h^2} \max_{t\geq 1} E \left\| \eta_t \eta_t^\top - \Omega(\tau_t) \right\| = O(\gamma_T).$$

We next apply the truncation method to investigate the rate of $I_{T,1}$. Define

$$u_t = \text{vec}\left[\eta_t \eta_t^\top - \Omega(\tau_t)\right],\ u_t' = u_t I(\|u_t\| \leq T^{2/\delta}),\ u_t'' = u_t - u_t'.$$

Since $\{u_t\}$ is a martingale difference sequence, we have

$$\begin{aligned}
I_{T,1} &= \max_{1\leq l\leq N_T} \left\| \frac{1}{T} \sum_{t=1}^{T} u_t' + u_t'' - E(u_t' + u_t'' \mid \mathcal{F}_{t-1}) K_h(\tau_t - s_l) \right\| \\
&\leq \max_{1\leq l\leq N_T} \left\| \frac{1}{T} \sum_{t=1}^{T} \left(u_t' - E(u_t' \mid \mathcal{F}_{t-1})\right) K_h(\tau_t - s_l) \right\| + \max_{1\leq l\leq N_T} \left\| \frac{1}{T} \sum_{t=1}^{T} u_t'' K_h(\tau_t - s_l) \right\| \\
&\quad + \max_{1\leq l\leq N_T} \left\| \frac{1}{T} \sum_{t=1}^{T} E(u_t'' \mid \mathcal{F}_{t-1}) K_h(\tau_t - s_l) \right\| = I_{T,11} + I_{T,12} + I_{T,13},
\end{aligned}$$

where the definitions of $I_{T,11}$, $I_{T,12}$, and $I_{T,13}$ should be obvious.

Let $d_T = \max_{1\leq t\leq T} \frac{1}{T} K_h(\tau_t - s_l)$. For $I_{T,12}$, by Hölder's inequality and Markov inequality, we have

$$\begin{aligned}
E|I_{T,12}| &\leq d_T \sum_{t=1}^{T} E\left\|u_t''\right\| \leq d_T \sum_{t=1}^{T} \left(E\|u_t\|^{\delta/2}\right)^{2/\delta} \left(\text{pr}\left(\|u_t\| > T^{2/\delta}\right)\right)^{1-2/\delta} \\
&\leq d_T \sum_{t=1}^{T} \left(E\|u_t\|^{\delta/2}\right)^{2/\delta} \left(\frac{E\|u_t\|^{\delta/2}}{T}\right)^{1-2/\delta} = O\left(T^{2/\delta}/(Th)\right) = o\left(\sqrt{\frac{\log T}{Th}}\right).
\end{aligned}$$

Similarly, we have $I_{T,13} = O_P(T^{2/\delta}/(Th)) = o\left(\sqrt{\frac{\log T}{Th}}\right)$. Then we just need to concentrate on $I_{T,11}$. For any $1 \leq l \leq N_T$, let $U_t = \frac{1}{T}(u_t' - E(u_t' \mid \mathcal{F}_{t-1}))K_h(\tau_t - s_l)$, then we have $E\left(U_t \mid \mathcal{F}_{t-1}\right) = 0$ and $\|U_t\| \leq 2T^{2/\delta} d_T$. Since $\sup_{t\geq 1} E\left(\|\eta_t\|^4 \mid \mathcal{F}_{t-1}\right) < \infty$ almost surely, we have

$$\begin{aligned}
&\max_{1\leq l\leq N_T} \left\| \sum_{t=1}^{T} E(U_t U_t^\top \mid \mathcal{F}_{t-1}) \right\| \leq 4 \max_{1\leq l\leq N_T} \sum_{t=1}^{T} E\left(\|u_t\|^2 \mid \mathcal{F}_{t-1}\right) \frac{K_h(\tau_t - s_l)^2}{T^2} \\
&\leq 4 d_T \sup_{t\geq 1} E\left(\|u_t\|^2 \mid \mathcal{F}_{t-1}\right) \max_{1\leq l\leq N_T} \frac{1}{T} \sum_{t=1}^{T} K_h(\tau_t - s_l) = O_{a.s.}\left(d_T\right).
\end{aligned}$$

Applying Lemma S5, if $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \to \infty$, we have

$$\begin{aligned}
\text{pr}\left(I_{T,11} > \sqrt{8M}\gamma_T\right) &\leq \text{pr}\left(I_{T,11} > \sqrt{8M}\gamma_T, \max_{1\leq l\leq N_T} \left\| \sum_{t=1}^{T} E(U_t U_t^\top \mid \mathcal{F}_{t-1}) \right\| \leq \frac{M}{Th}\right) \\
&\leq N_T \exp\left(-\frac{8M\gamma_T^2}{2(\frac{M}{Th} + \gamma_T 2T^{\frac{2}{\delta}} d_T)}\right) \leq N_T \exp\left(-4\log(T)\right) = o(1).
\end{aligned}$$

The proof is now completed. □

Proof of Lemma S7. (1). By Lemma S6,

$$SX_C = \begin{pmatrix} (X_{\widetilde{C},1}^\top, 0_{d\times(l-s)})(X_{\widetilde{C},\tau_1}^\top K_{\tau_1} X_{\widetilde{C},\tau_1})^{-1} X_{\widetilde{C},\tau_1}^\top K_{\tau_1} X_C \\ \vdots \\ (X_{\widetilde{C},T}^\top, 0_{d\times(l-s)})(X_{\widetilde{C},\tau_T}^\top K_{\tau_T} X_{\widetilde{C},\tau_T})^{-1} X_{\widetilde{C},\tau_T}^\top K_{\tau_T} X_C \end{pmatrix}$$
$$= \begin{pmatrix} X_{\widetilde{C},1}^\top \Sigma_{X_{\widetilde{C}}}^{-1}(\tau_1) \Sigma_{X_{C,\widetilde{C}}}^\top(\tau_1) \\ \vdots \\ X_{\widetilde{C},T}^\top \Sigma_{X_{\widetilde{C}}}^{-1}(\tau_T) \Sigma_{X_{C,\widetilde{C}}}^\top(\tau_T) \end{pmatrix} (1 + o_P(1)),$$

which follows that

$$\frac{1}{T} X_C^\top (I_{dT} - S)^\top (I_{dT} - S) X_C$$
$$= \frac{1}{T}\sum_{t=1}^{T} \left(X_{C,t}^\top - X_{\widetilde{C},t}^\top \Sigma_{X_{\widetilde{C}}}^{-1}(\tau_t) \Sigma_{X_{C,\widetilde{C}}}^\top(\tau_t) \right)^\top \left(X_{C,t}^\top - X_{\widetilde{C},t}^\top \Sigma_{X_{\widetilde{C}}}^{-1}(\tau_t) \Sigma_{X_{C,\widetilde{C}}}^\top(\tau_t) \right) + o_P(1).$$

Note that each element of $\Sigma_{X_{\widetilde{C}}}(\tau)$ and $\Sigma_{X_{C,\widetilde{C}}}(\tau)$ is Lipschitz continuous. Thus, by Lemma S4 and S1, the result holds.

(2). Let $\rho_T = h^2 + \sqrt{\log T/(Th)}$. By Lemma S6.1, we have

$$[X_{\widetilde{C},t}^\top, 0_{d\times(l-s)}](X_{\widetilde{C},\tau_t}^\top K_{\tau_t} X_{\widetilde{C},\tau_t})^{-1} X_{\widetilde{C},\tau_t}^\top K_{\tau_t} \widetilde{X} = X_{\widetilde{C},t}^\top \theta(\tau_t)(1 + O_P(\rho_T))$$

uniformly over $1 \le t \le T$. Hence, we have

$$X_C^\top (I_{dT} - S)^\top (I_{dT} - S)\widetilde{X}$$
$$= \sum_{t=1}^{T} \left(X_{C,t} X_{\widetilde{C},t}^\top - \Sigma_{X_{C,\widetilde{C}}}(\tau_t) \Sigma_{X_{\widetilde{C}}}^{-1}(\tau_t) X_{\widetilde{C},t} X_{\widetilde{C},t}^\top (1 + O_P(c_T)) \right) \theta(\tau_t) \cdot O_P(c_T)$$
$$= O_P(T\rho_T^2),$$

where the last equality follows from Lemma S4. Finally, the result holds since $O_P(T\rho_T^2) = o_P(\sqrt{T})$ by Assumption 7. □

Proof of Lemma S8. (1). Write

$$\sigma_T^2 = \frac{1}{T^2 h}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left[\int_{-1}^{1} K(u) K\left(u + \frac{t-s}{Th}\right) du\right]^2$$
$$= \frac{1}{T^2 h}\sum_{t=2}^{T}\sum_{j=1}^{t-1}\left[\int_{-1}^{1} K(u) K\left(u + \frac{j}{Th}\right) du\right]^2$$
$$= \frac{1}{Th}\sum_{j=1}^{T-1}(1 - j/T)\left[\int_{-1}^{1} K(u) K\left(u + \frac{j}{Th}\right) du\right]^2$$
$$= \int_0^\infty (1 - vh)\left[\int_{-1}^{1} K(u) K(u+v)\, du\right]^2 dv + O(1/(Th))$$

$$\to \int_0^2 \left[\int_{-1}^{1-v} K(u) K(u+v) du \right]^2 dv.$$

(2). Write

$$\begin{aligned}
\max_t |a_t| &= \max_t \left| \sum_{i=1}^{t-1} \frac{1}{T^2 h} \left[\int_{-1}^{1} K(u) K\left(u + \frac{i}{Th}\right) du \right]^2 \right| \\
&\leq \frac{1}{T^2 h} \sum_{i=1}^{T} \left[\int_{-1}^{1} K(u) K\left(u + \frac{i}{Th}\right) du \right]^2 \\
&= \frac{1}{T} \int_0^{\infty} \left[\int_{-1}^{1} K(u) K(u+v) du \right]^2 \mathrm{d}v (1 + o(1)) = O(1/T).
\end{aligned}$$

(3). Write

$$\begin{aligned}
\sum_{s=1}^{T-J} w_{s,s+J}^2 &= \sum_{s=1}^{T-J} \frac{1}{T^2 h} \left[\int_{-1}^{1} K(u) K\left(u + \frac{J}{Th}\right) du \right]^2 \\
&= \frac{T-J}{T^2 h} \left[\int_{-1}^{1} K(u) K\left(u + \frac{J}{Th}\right) du \right]^2 = O(1/(Th)).
\end{aligned}$$

(4). Write

$$\begin{aligned}
& T \sum_{s=1}^{T-1} b_s^2 \\
&= \frac{1}{T^3 h^2} \sum_{j=1}^{T-1} \sum_{t=1+j}^{T} \sum_{s=1+j}^{T} \left[\int_{-1}^{1} K(u) K\left(u + \frac{t-j}{Th}\right) \mathrm{d}u \right]^2 \left[\int_{-1}^{1} K(u) K\left(u + \frac{s-j}{Th}\right) du \right]^2 \\
&= \frac{1}{T^3 h^2} \sum_{j=1}^{T-1} \sum_{i=1}^{T-j} \sum_{k=1}^{T-j} \left[\int_{-1}^{1} K(u) K\left(u + \frac{i}{Th}\right) \mathrm{d}u \right]^2 \left[\int_{-1}^{1} K(u) K\left(u + \frac{k}{Th}\right) du \right]^2 \\
&\leq \frac{1}{T^3 h^2} \sum_{j=1}^{T} \sum_{i=1}^{T} \sum_{k=1}^{T} \left[\int_{-1}^{1} K(u) K\left(u + \frac{i}{Th}\right) \mathrm{d}u \right]^2 \left[\int_{-1}^{1} K(u) K\left(u + \frac{k}{Th}\right) du \right]^2 \\
&\simeq \left(\frac{1}{Th} \sum_{i=1}^{T} \left[\int_{-1}^{1} K(u) K\left(u + \frac{i}{Th}\right) du \right]^2 \right)^2 = O(1).
\end{aligned}$$

(5). By Cauchy-Schwarz inequality,

$$\begin{aligned}
& \sum_{k=1}^{T-1} \sum_{t=1}^{k-1} \left[\sum_{j=k+1}^{T} w_{k,j} w_{t,j} \right]^2 \leq \frac{1}{T^4 h^2} \sum_{k=1}^{T-1} \sum_{t=1}^{k-1} \Bigg\{ \left(\sum_{j=1+k}^{T} \left[\int_{-1}^{1} K(u) K\left(u + \frac{j-k}{Th}\right) du \right]^2 \right) \\
& \cdot \left(\sum_{j=1+k}^{T} \left[\int_{-1}^{1} K(u) K\left(u + \frac{j-t}{Th}\right) du \right]^2 \right) \Bigg\} \\
&\leq \frac{M}{T^3 h} \sum_{k=1}^{T-1} \sum_{t=1}^{k-1} \sum_{j=1+k}^{T} \left[\int_{-1}^{1} K(u) K\left(u + \frac{j-t}{Th}\right) du \right]^2 \\
&\leq \frac{M}{T^3 h} \sum_{k=1}^{T-1} \sum_{j=2}^{T} \sum_{t=1}^{j-1} \left[\int_{-1}^{1} K(u) K\left(u + \frac{j-t}{Th}\right) du \right]^2 = O(1/T).
\end{aligned}$$

The proof is now completed. □

Proof of S9. Note that $\xi_t = (\xi_{t,1}, ..., \xi_{t,d})^\top$, and we first prove

$$\left\|\left\|\sum_{t=1}^{T} \xi_{t,i}\right\|\right\|_p^{p^*} \leq O(1) \sum_{t=1}^{T} \|\|\xi_{t,i}\|\|_p^{p^*}$$

for $1 \leq i \leq d$. If $p \geq 2$, then $p^* = 2$, by Burkholder inequality (e.g., Theorem 2.10 of Hall and Heyde, 1980) and Minkowski inequality, we have

$$\begin{aligned}\left\|\left\|\sum_{t=1}^{T} \xi_{t,i}\right\|\right\|_p^{2} &\leq \left\{O(1) \cdot E\left[\left(\sum_{t=1}^{T} |\xi_{t,i}|^2\right)^{p/2}\right]\right\}^{2/p} \\ &\leq O(1) \sum_{t=1}^{T} \left(E\left[|\xi_{t,i}|^p\right]\right)^{2/p} = O(1) \sum_{t=1}^{T} \|\|\xi_{t,i}\|\|_p^2.\end{aligned}$$

If $1 < p < 2$, then $0 < \frac{p}{2} < 1$, by Burkholder inequality and $|\sum_{i=1}^d a_i|^r \leq \sum_{i=1}^d |a_i|^r$ for $r \in (0, 1]$, we have

$$\begin{aligned}\left\|\left\|\sum_{t=1}^{T} \xi_{t,i}\right\|\right\|_p^{p^*} &\leq \left\{O(1) \cdot E\left[\left(\sum_{t=1}^{T} |\xi_{t,i}|^2\right)^{p/2}\right]\right\}^{p^*/p} \\ &\leq \left\{O(1) \cdot E\left[\sum_{t=1}^{T} |\xi_{t,i}|^p\right]\right\}^{p^*/p} = O(1) \cdot \left\{\sum_{t=1}^{T} E|\xi_{t,i}|^p\right\}^{p^*/p} \\ &\leq O(1) \sum_{t=1}^{T} \left(E\left[|\xi_{t,i}|^p\right]\right)^{p^*/p} = O(1) \sum_{t=1}^{T} \|\|\xi_{t,i}\|\|_p^{p^*}.\end{aligned}$$

In addition, for fixed d, since $|\sum_{i=1}^d a_i|^r \leq \sum_{i=1}^d |a_i|^r$ for $r \in (0, 1]$, and $|\sum_{i=1}^d a_i|^r \leq d^{r-1} \sum_{i=1}^d |a_i|^r$ for $r > 1$, we have

$$\begin{aligned}\left\|\left\|\sum_{t=1}^{T} \xi_t\right\|\right\|_p^{p^*} &= \left\{E\left[\left(\sum_{i=1}^{d} \xi_{\cdot,i}^2\right)^{p/2}\right]\right\}^{p^*/p} \leq M\left\{\sum_{i=1}^{d} E\left|\xi_{\cdot,i}\right|^p\right\}^{p^*/p} \\ &\leq M \sum_{i=1}^{d} \left\|\left\|\sum_{t=1}^{T} \xi_{t,i}\right\|\right\|_p^{p^*} \leq M \sum_{t=1}^{T} \sum_{i=1}^{d} \|\|\xi_{t,i}\|\|_p^{p^*} = M \sum_{t=1}^{T} \sum_{i=1}^{d} \{E|\xi_{t,i}|^p\}^{p^*/p} \\ &= M \sum_{t=1}^{T} \left\{\sum_{i=1}^{d} \{E|\xi_{t,i}|^p\}^{p^*/p}\right\}^{p/p^* \times p^*/p} \leq M \sum_{t=1}^{T} \left\{E\left(\sum_{i=1}^{d} |\xi_{t,i}|^p\right)\right\}^{p^*/p} \\ &= M \sum_{t=1}^{T} \left\{E\left(\sum_{i=1}^{d} |\xi_{t,i}|^p\right)^{\frac{2}{p} \times \frac{p}{2}}\right\}^{p^*/p} \leq M \sum_{t=1}^{T} \left\{E\left(\sum_{i=1}^{d} |\xi_{t,i}|^2\right)^{\frac{p}{2}}\right\}^{p^*/p} \\ &= M \sum_{t=1}^{T} \|\|\xi_t\|\|_p^{p^*},\end{aligned}$$

where $\xi_{\cdot,i} = \sum_{t=1}^T \xi_{t,i}$. The proof is now completed. □

Proof of S10. The proof of this lemma is a modified version of the martingale approximation method in Wu and Shao (2007). Without loss of generality, let $E(w_t^*) = 0$. For any integer $I \geq 1$ we introduce the truncated process $z_{t-1,I}^* = E\left(z_{t-1}^* \mid \mathcal{F}_{t-I}\right)$. Then $z_{t-1,I}^* = 0$ if $t \leq I$ and $z_{t-1,I}^* = \sum_{s=1}^{t-I} w_{s,t} y_s^*$ for $1 \leq I < t$. For $2 \leq t \leq T$, by Lemma S9,

$$\left|\left|\left| z_{t-1,I}^* - z_{t-1}^* \right|\right|\right|_\delta^2 \leq M \max_t |||y_t^*|||_\delta^2 \sum_{s=\max(1,t-I+1)}^{t-1} w_{s,t}^2 = O\left(\sum_{s=\max(1,t-I+1)}^{t-1} w_{s,t}^2 \right).$$

Let $L(I) = \sum_{J=1}^{I} l(J)$ and $V(I) = \sum_{t=2}^{T} \operatorname{tr}(w_t^* z_{t-1,I}^* z_{t-1,I}^{*,\top})$, where $l(J) = \sum_{s=1}^{T-J} w_{s,s+J}^2$ and $T(I) = \sum_{t=2}^{T} \operatorname{tr}\left[E(w_t^* \mid \mathcal{F}_{t-I}) z_{t-1,I}^* z_{t-1,I}^{*,\top} \right]$. By Cauchy-Schwarz inequality and Lemma S8 (iii), if $I/(Th) \to 0$, we have

$$\begin{aligned}
E|V(1) - V(I)| &\leq \sum_{t=2}^{T} E\left| \operatorname{tr}\left[w_t^* \left(z_{t-1}^* z_{t-1}^{*,\top} - z_{t-1,I}^* z_{t-1,I}^{*,\top} \right) \right] \right| \\
&\leq \sum_{t=2}^{T} |||w_t^*||| \cdot \left|\left|\left| z_{t-1}^* - z_{t-1,I}^* \right|\right|\right|_4 \left|\left|\left| z_{t-1}^* + z_{t-1,I}^* \right|\right|\right|_4 \leq M \sum_{t=2}^{T} \left|\left|\left| z_{t-1}^* - z_{t-1,I}^* \right|\right|\right|_4 a_t^{1/2} \\
&\leq M \left\{ \sum_{t=2}^{T} \left|\left|\left| z_{t-1}^* - z_{t-1,I}^* \right|\right|\right|_4^2 \right\}^{1/2} \left\{ \sum_{t=2}^{T} a_t \right\}^{1/2} = O(1)[L(I)]^{1/2} \to 0,
\end{aligned}$$

where we have used the fact that

$$\begin{aligned}
\left|\left|\left| z_{t-1}^* + z_{t-1,I}^* \right|\right|\right|_4 &\leq \left\{ M \sum_{s=1}^{t-1} |||w_{s,t} y_s^*|||_4^2 \right\}^{1/2} + \left\{ M \sum_{s=1}^{t-I} |||w_{s,t} y_s^*|||_4^2 \right\}^{1/2} \\
&= O\left(\left\{ \sum_{s=1}^{t-1} w_{s,t}^2 \right\}^{1/2} \right) = o(a_t^{1/2}).
\end{aligned}$$

Define the projection operator $\mathcal{P}_t \xi = E(\xi \mid \mathcal{F}_t) - E(\xi \mid \mathcal{F}_{t-1})$. For $0 \leq j \leq I - 1$, let $U(j) = \sum_{t=2}^{T} \operatorname{tr}\left[\left(\mathcal{P}_{t-j} w_t^* \right) z_{t-1,I}^* z_{t-1,I}^{*,\top} \right]$, then

$$V(I) - T(I) = \sum_{t=2}^{T} \operatorname{tr}\left[\left(\sum_{j=0}^{I-1} \mathcal{P}_{t-j} w_t^* \right) z_{t-1,I}^* z_{t-1,I}^{*,\top} \right] = \sum_{j=0}^{I-1} U(j).$$

Note that $\left\{ \left(\mathcal{P}_{t-j} w_t^* \right) z_{t-1,I}^* z_{t-1,I}^{*,\top} \right\}_{t=2}^{T}$ is a martingale difference sequence since

$$E\left\{ \left(\mathcal{P}_{t-j} w_t^* \right) z_{t-1,I}^* z_{t-1,I}^{*,\top} \mid \mathcal{F}_{t-j-1} \right\} = \left[E(w_t^* \mid \mathcal{F}_{t-j-1}) - E(w_t^* \mid \mathcal{F}_{t-j-1}) \right] z_{t-1,I}^* z_{t-1,I}^{*,\top} = 0.$$

By Lemma S8 (ii), Lemma S9 and Cauchy-Schwarz inequality, since $|||\mathcal{P}_{t-j} w_t^*|||_{\delta/2} \leq 2|||w_t^*|||_{\delta/2} < \infty$,

$$\begin{aligned}
|||U(j)|||_{\delta/4}^{\delta/4} &\leq M \sum_{t=2}^{T} \left|\left|\left| \left(\mathcal{P}_{t-j} w_t^* \right) z_{t-1,I}^* z_{t-1,I}^{*,\top} \right|\right|\right|_{\delta/4}^{\delta/4} \leq M \sum_{t=2}^{T} \left|\left|\left| z_{t-1,I}^* \right|\right|\right|_\delta^{\delta/2} \\
&\leq M \sum_{t=2}^{T} a_t^{\delta/4} \leq M \max_t a_t^{\delta/4-1} \sum_{t=2}^{T} a_t = O\left(T^{1-\delta/4} \right).
\end{aligned}$$

In addition, by $E|V(1)-V(I)| \to 0$,

$$E|V(1)| \leq \|V(I)-T(I)\|_{\delta/4} + E|T(I)| + o(1)$$
$$\leq \sum_{j=0}^{I-1} \|U(j)\|_{\delta/4} + \max_t \|E(w_t^*|\mathcal{F}_{t-I})\| \sum_{t=2}^{T} \left\| z_{t-1,I}^* \right\|_4^2 = o(1),$$

since $\max_t \|E(w_t^* \mid \mathcal{F}_{t-I})\| \to 0$ as $I \to \infty$. The proof is now complete. □

Proof of S11. For notational simplicity, let $H_t = I_{d^2}$. Write

$$\sum_{t=2}^{T} \operatorname{tr}\left[z_{t-1}^* z_{t-1}^{*,\top} - E\left(z_{t-1}^* z_{t-1}^{*,\top}\right)\right]$$
$$= \sum_{t=2}^{T}\sum_{s=1}^{t-1} \operatorname{tr}\left[\left(Z_s\eta_s\eta_s^\top Z_s^\top - E\left(Z_s\eta_s\eta_s^\top Z_s^\top\right)\right) w_{s,t}^2\right]$$
$$+2\sum_{t=3}^{T}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1} \operatorname{tr}\left[Z_{s_1}\eta_{s_1}\eta_{s_2} Z_{s_2}^\top w_{s_1,t} w_{s_2,t}\right] = I_{T,1} + 2I_{T,2},$$

where the definitions of $I_{T,1}$ and $I_{T,2}$ should be obvious.

Consider $I_{T,1}$. Write

$$I_{T,1} = \sum_{t=2}^{T}\sum_{s=1}^{t-1} \operatorname{tr}\left[\left(\eta_s\eta_s^\top - \Omega(\tau_s)\right) Z_s^\top Z_s\right] w_{s,t}^2 + \sum_{t=2}^{T}\sum_{s=1}^{t-1} \operatorname{tr}\left[\Omega(\tau_s)\left(Z_s^\top Z_s - E\left(Z_s^\top Z_s\right)\right)\right] w_{s,t}^2$$
$$= \frac{1}{T}\sum_{s=1}^{T-1} \operatorname{tr}\left[\left(\eta_s\eta_s^\top - \Omega(\tau_s)\right) Z_s^\top Z_s\right]\left(T\sum_{t=s+1}^{T} w_{s,t}^2\right)$$
$$+\frac{1}{T}\sum_{s=1}^{T-1} \operatorname{tr}\left[\Omega(\tau_s)\left(Z_s^\top Z_s - E\left(Z_s^\top Z_s\right)\right)\right]\left(T\sum_{t=s+1}^{T} w_{s,t}^2\right) = I_{T,11} + I_{T,12}.$$

Since $\operatorname{tr}\left[\left(\eta_s\eta_s^\top - \Omega(\tau_s)\right) Z_s^\top Z_s\right]$, $s = 1, 2, \ldots$ are martingale differences and $T\sum_{t=1}^{T} w_{s,t}^2 = O(1)$ by Lemma S8.2, we have $I_{T,11} = o_P(1)$. In addition, by Lemma S4, we have $I_{T,12} = o_P(1)$.

Next, consider $I_{T,2}$. By Lemma S9, Cauchy-Schwarz inequality and Lemma S8.5,

$$\|I_{T,2}\|^2 \leq M \sum_{s_1=2}^{T-1} \left\|\operatorname{tr}\left[Z_{s_1}\eta_{s_1}\sum_{s_2=1}^{s_1-1}\eta_{s_2} Z_{s_2}^\top \sum_{t=s_1+1}^{T} w_{s_1,t}w_{s_2,t}\right]\right\|^2$$
$$\leq M \sum_{s_1=2}^{T-1} \|Z_{s_1}\eta_{s_1}\|_4^2 \left\|\sum_{s_2=1}^{s_1-1} Z_{s_2}\eta_{s_2}\sum_{t=s_1+1}^{T} w_{s_1,t}w_{s_2,t}\right\|_4^2$$
$$\leq M \sum_{s_1=2}^{T-1} \|Z_{s_1}\eta_{s_1}\|_4^2 \sum_{s_2=1}^{s_1-1} \left\|Z_{s_2}\eta_{s_2}\sum_{t=s_1+1}^{T} w_{s_1,t}w_{s_2,t}\right\|_4^2$$
$$= O\left(\sum_{s_1=2}^{T-1}\sum_{s_2=1}^{s_1-1}\Big(\sum_{t=s_1+1}^{T} w_{s_1,t}w_{s_2,t}\Big)^2\right) = O(1/T).$$

Combining the above results, the proof is now completed. □

Proof of S12. Note that $y_t^{*,\top} H_t z_{t-1}^*$, $t \in \mathbb{Z}$, are martingale differences with respect to the filtration $\mathcal{F}_t$.

We apply the martingale central limit theory to prove the asymptotic normality of Q_T. By Assumption 7,

$$E\left[\|y_t^*\|^\delta\right] \le E\left[E\left(\|\epsilon_t\|^\delta \mid \mathcal{F}_{t-1}\right)\|z_{t-1}\otimes I_d\|^\delta\right] < E\left[O(1)\,\|z_{t-1}\|^\delta\right] < \infty.$$

By the Cauchy-Schwarz inequality, Lemma S9 and Lemma S8.2, Lindeberg conditions are satisfied since

$$\begin{aligned}
&\sum_{t=2}^{T} \left\|\left\|y_t^* H_t z_{t-1}^*\right\|\right\|_{\delta/2}^{\delta/2} \le \sum_{t=2}^{T} \left\|\left\|y_t^* H_t\right\|\right\|_{\delta}^{\delta/2} \left\|\left\|z_{t-1}^*\right\|\right\|_{\delta}^{\delta/2} \\
&\le M \max_t \left\|\left\|y_t^*\right\|\right\|_\delta^\delta \sum_{t=2}^{T} a_t^{\delta/4} = O(1)\cdot \max_t a_t^{\delta/4-1} = o(1).
\end{aligned}$$

Applying Lemmas S10 and S11 with $w_t^* = E\left(H_t^\top y_t^* y_t^{*,\top} H_t \mid \mathcal{F}_{t-1}\right)$, we then have the convergence of conditional variance

$$\begin{aligned}
&\sum_{t=2}^{T} \operatorname{tr}\left[E\left(H_t^\top y_t^* y_t^{*,\top} H_t \mid \mathcal{F}_{t-1}\right) z_{t-1}^* z_{t-1}^{*,\top}\right] \\
\to_P &\sum_{t=2}^{T} \operatorname{tr}\left[E\left(H_t^\top y_t^* y_t^{*,\top} H_t\right) E\left(z_{t-1}^* z_{t-1}^{*,\top}\right)\right].
\end{aligned}$$

By the martingale central limit theory, the proof is now completed.