

Department of Econometrics and Business Statistics

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

On Time-Varying VAR models: Estimation, Testing and Impulse Response Analysis

Yayi Yan, Jiti Gao and Bin Peng

November 2021

Working Paper 17/21

On Time-Varying VAR Models: Estimation, Testing and Impulse Response Analysis

YAYI YAN, JITI GAO AND BIN PENG

Monash University, Melbourne, Australia

October 31, 2021

Abstract

Vector autoregressive (VAR) models are widely used in practical studies, e.g., forecasting, modelling policy transmission mechanism, and measuring connection of economic agents. To better capture the dynamics, this paper introduces a new class of time-varying VAR models in which the coefficients and covariance matrix of the error innovations are allowed to change smoothly over time. Accordingly, we establish a set of theories, including the impulse responses analyses subject to both of the short-run timing and the long-run restrictions, an information criterion to select the optimal lag, and a Wald-type test to determine the constant coefficients. Simulation studies are conducted to evaluate the theoretical findings. Finally, we demonstrate the empirical relevance and usefulness of the proposed methods through an application to the transmission mechanism of U.S. monetary policy.

Keywords: Multivariate Dynamic Time Series; Time-Varying Impulse Response; Testing for Parameter Stability

JEL Classification: C14, C32, E52

Corresponding author: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Email: Jiti.Gao@monash.edu. The authors of this paper would like to thank George Athanasopoulos, Rainer Dahlhaus, David Frazier, Oliver Linton, Gael Martin, Peter CB Phillips and Wei Biao Wu 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.

1 Introduction

Vector autoregressive (VAR) models as well as their extensions are among some of the most popular frameworks for modelling dynamic interactions among multiple variables. These models arise mainly as a response to the “incredible” identification conditions embedded in the large-scale macroeconomic models (Sims, 1980). VAR modelling begins with minimal restrictions on the multivariate dynamic models. Gradually armed with identification information, VAR models and their statistical tool-kits like impulse response functions become powerful tools for conducting policy analysis. We refer interested readers to Stock and Watson (2001) for a comprehensive review. Despite the popularity, linear VAR models can always be rejected by data in empirical studies (Tsay, 1998). For example, Stock and Watson (2016) point out, “*changes associated with the Great Moderation go beyond reduction in variances to include changes in dynamics and reduction in predictability*”.

To go beyond linear VAR models, various parametric time-varying VAR models have been proposed (e.g., Tsay, 1998, Sims and Zha, 2006, and the references therein) in order to allow for certain changes in economic relationship. However, model misspecification and parameter instability may undermine the performance of the proposed parametric models. As pointed out by Hansen (2001), it may seem unlikely that structural change could be immediate and might seem more reasonable to allow structural change to take a period of time to take effect. Practically, it may be more reasonable to allow smooth structural changes over a period of time rather than in an abrupt manner. To model the dynamic transit, an important strand of the VAR literature assumes that the coefficients of VAR models evolve in a random way (e.g., Primiceri, 2005; Petrova, 2019), and the estimation procedure relies on extensive Markov Chain Monte Carlo (MCMC) draws plus the use of a variety of filters, such as Kalman, or related filters. However, by doing so, the asymptotic properties of the estimated model coefficients as well as the corresponding impulse responses are still unclear (Giraitis et al., 2014).

Meanwhile, there is a separate literature about using nonparametric methods to estimate deterministically unknown time-varying parameters for autoregressive models. Up to this point, it is worth bringing up the terminology “local stationarity”, as a bulk of the literature is carried out using the local stationarity technique, which at least dates back to the seminal work by Dahlhaus (1996). Recent developments focus on univariate autoregressive models (Dahlhaus and Rao, 2006; Zhang and Wu, 2012; Richter and Dahlhaus, 2019; Sun et al., 2021). To the best of our knowledge, it has had very little success to extend the local stationarity technique to multivariate settings, such as VAR models with deterministic time-varying parameters. In some specific cases where different locally stationary univariate time series may be approximated by their stationary versions on the same segments, the local stationarity technique may be applicable to such specific multivariate settings. However, univariate time series of general multivariate time series may have quite different patterns and behaviours, such as the three univariate time series plotted in Figure 1 of Section 5.

To address the aforementioned issues, this paper therefore proposes a class of deterministic time-varying VAR models where both VAR coefficients and covariance matrix of the model’s error innovations are allowed to change smoothly over time. We develop a time-varying vector moving

average infinity ($\text{VMA}(\infty)$) representation for a class of VAR models before we are able to establish uniform consistency results and a joint central limit theory for kernel-based estimators of both VAR coefficients and covariance matrix, which facilitate the inference on time-varying structural impulse responses. This type of impulse responses is of major importance in typical VAR applications (Inoue and Kilian, 2013, 2020; Paul, 2020). In addition, inferences on structural impulse responses subject to both short-run timing and long-run restrictions are considered, so that the proposed model can better capture the simultaneous relations among multiple variables of interest over time. Such modelling strategy is especially useful for analyzing multivariate time series over a long horizon, since it offers a comprehensive treatment on tracking interests which are affected by frequently updated policies, environment, system, etc. In an economy system consisting of inflation, unemployment and interest rates, we discuss time-varying impacts of the interest rate change, which helps stabilize fluctuations in inflation and unemployment in the long-run. Under the proposed framework, it is achieved by investigating the corresponding time-varying impulse response functions.

We now comment on the literature closely related to the parameter stability testing problem we are considering in this paper. Detecting and estimating parametric components in univariate time-varying autoregressive models has been studied by Zhang and Wu (2012). Recently, Truquet (2017) considers parameter stability testing for univariate time-varying autoregressive conditional heteroscedasticity. More recently, Lütkepohl and Schlaak (2021) discuss testing for time-varying impact effects through a change-point mechanism. We develop a simple test for checking whether some of the time-varying coefficients (if not all) reduce to constant coefficients involved in the VAR models. It can be used to test whether the policy transmission mechanism is changing over time, which is of great importance in the macroeconomic literature (e.g., Primiceri, 2005; Paul, 2020). To give another example, consider our empirical study on the ongoing debate about the high inflation during 1970-1980, i.e., whether the high inflation is due to bad monetary policy (Primiceri, 2005; Sims and Zha, 2006). Mathematically, the discrepancy comes down to specification testing on the coefficient matrices of the VAR models by the proposed test.

In summary, our contributions are in three-fold. First, we propose a class of time-varying $\text{VAR}(p)$ models, and develop a time-varying $\text{VMA}(\infty)$ representation for the $\text{VAR}(p)$ models before we establish an estimation theory for time-varying structural impulse responses, which are often of great interest to describe how the economy reacts over time to structural shocks. The proposed estimation is conducted subject to both short-run timing and long-run restrictions, which have attracted considerable attention in the literature (Kilian and Lütkepohl, 2017). Second, we develop a Wald-type test statistic for detecting time-invariant parameters in time-varying VAR models before we show that the proposed test statistic is asymptotically normally distributed under both the null hypothesis and a sequence of local alternatives. Third, in an empirical study, we investigate the changing dynamics of three key U.S. macroeconomic variables (inflation, unemployment, and interest rate), and uncover a fall in the volatilities of exogenous shocks. We observe that there exists a substantial time-variation in the policy transmission mechanism, and the “price puzzle” is limited to periods of bad monetary policy.

The organization of this paper is as follows. Section 2 considers a class of time-varying VAR models, and establish asymptotic properties for the estimated time-varying impulse response functions. Section 3 describes the proposed test and establishes the corresponding asymptotic theory. Section 4 discusses some implementation issues and presents comprehensive simulation studies. Section 5 provides a case study to demonstrate the empirical relevance of the proposed models and estimation theory. Section 6 concludes. The proofs of the main results are given in Appendix A. Preliminary lemmas and their proofs are given in Appendix B.

Before proceeding further, it is convenient to introduce some notations: $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix; \otimes denotes the Kronecker product; \mathbf{I}_a stands for an $a \times a$ identity matrix; $\mathbf{0}_{a \times b}$ stands for an $a \times b$ matrix of zeros, and we write $\mathbf{0}_a$ for short when $a = b$; for a function $g(w)$, let $g^{(j)}(w)$ be the j^{th} derivative of $g(w)$, where $j \geq 0$ 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; let $\tilde{c}_k = \int_{-1}^1 u^k K(u) du$ and $\tilde{v}_k = \int_{-1}^1 u^k K^2(u) du$ for integer $k \geq 0$; $\text{vec}(\cdot)$ stacks the elements of an $m \times n$ matrix as an $mn \times 1$ vector; for any $a \times a$ square matrix \mathbf{A} , $\text{vech}(\mathbf{A})$ denotes the $\frac{1}{2}a(a+1) \times 1$ vector obtained from $\text{vec}(\mathbf{A})$ by eliminating all supra-diagonal elements of \mathbf{A} ; $\text{tr}(\mathbf{A})$ denotes the trace of \mathbf{A} . Finally, \rightarrow_P and \rightarrow_D denote convergence in probability and convergence in distribution, respectively.

2 The Time-Varying VAR(p) Model

Suppose that we observe $\{\mathbf{x}_{-p+1}, \dots, \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T\}$ from the following data generating process:

$$\mathbf{x}_t = \mathbf{a}(\tau_t) + \sum_{j=1}^p \mathbf{A}_j(\tau_t) \mathbf{x}_{t-j} + \boldsymbol{\eta}_t \quad \text{with} \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t) \boldsymbol{\epsilon}_t, \quad (2.1)$$

where $\tau_t = t/T$, $\mathbf{a}(\cdot)$ and $\mathbf{A}_j(\cdot)$ are respectively a $d \times 1$ vector and $d \times d$ matrices of unknown functional coefficients, and $\boldsymbol{\omega}(\tau)$ is a $d \times d$ matrix of unknown functions capturing the heteroskedasticity of the error components. Allowing $\boldsymbol{\omega}(\cdot)$ to vary over time is important theoretically and practically, because a constant covariance matrix implies that the shock to the i^{th} variable of \mathbf{x}_t has a time-invariant effect on the j^{th} variable of \mathbf{x}_t , restricting simultaneous interactions among multiple variables to be time-invariant (Primiceri, 2005). For the time being, we assume p is known, and shall come back to the choice of p in Section 2.3.

The following conditions are necessary for our development.

Assumption 1.

1. The roots of $\mathbf{I}_d - \mathbf{A}_1(\tau)L - \dots - \mathbf{A}_p(\tau)L^p = \mathbf{0}_d$ all lie outside the unit circle uniformly in $\tau \in [0, 1]$.
2. Each element of $\mathbf{A}(\tau) = [\mathbf{a}(\tau), \mathbf{A}_1(\tau), \dots, \mathbf{A}_p(\tau)]$ is second order continuously differentiable on $[0, 1]$ and $\mathbf{A}(\tau) = \mathbf{A}(0)$ for $\tau < 0$.

3. Each element of $\boldsymbol{\omega}(\tau)$ is second order continuously differentiable on $[0, 1]$. Moreover, $\boldsymbol{\Omega}(\tau) = \boldsymbol{\omega}(\tau)\boldsymbol{\omega}(\tau)^\top$ is positive definite uniformly in $\tau \in [0, 1]$ and $\boldsymbol{\omega}(\tau) = \boldsymbol{\omega}(0)$ for $\tau < 0$.

Assumption 2. $\{\boldsymbol{\epsilon}_t\}_{t=-\infty}^\infty$ is a martingale difference sequence (m.d.s.) adapted to the filtration $\{\mathcal{F}_t\}$, where $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \dots)$ is the σ -field generated by $(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \dots)$, $E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top | \mathcal{F}_{t-1}] = \mathbf{I}_d$ almost surely (a.s.), and $\max_t E \|\boldsymbol{\epsilon}_t\|^\delta < \infty$ for some $\delta > 4$.

Assumption 1.1 ensures the eigenvalues of the companion matrix $\boldsymbol{\Phi}(\tau)$ of (2.3) below all lie inside the unit circle uniformly over $\tau \in [0, 1]$. As a consequence, (2.1) cannot include any unit-root or explosive components. Similar treatments have also been adopted to investigate univariate locally stationary models in the literature (e.g., Assumption T3 of Zhang and Wu, 2012). Assumptions 1.2 and 1.3 allow the underlying data generating process to evolve over time in a smooth manner. The conditions $\mathbf{A}(\tau) = \mathbf{A}(0)$ and $\boldsymbol{\omega}(\tau) = \boldsymbol{\omega}(0)$ for $\tau < 0$ gives

$$\mathbf{x}_t = \mathbf{a}(0) + \sum_{j=1}^p \mathbf{A}_j(0) \mathbf{x}_{t-j} + \boldsymbol{\omega}(0) \boldsymbol{\epsilon}_t, \quad (2.2)$$

which basically assumes that \mathbf{x}_t behaves like a parametric VAR(p) model for $t \leq 0$. A similar condition can be found in Vogt (2012) for a nonparametric time-varying time series model.

Assumption 2 imposes some conditions on the innovation error terms, and are standard in the VAR literature (cf., Lütkepohl, 2005).

With these conditions, the following proposition says (2.1) admits a time-varying vector moving average infinity (VMA(∞)) representation, which sheds a light on how to recover the time-varying structural impulse responses.

Proposition 2.1. *Under Assumptions 1 and 2, there exists a time-varying VMA(∞) process of the form:*

$$\tilde{\mathbf{x}}_t = \boldsymbol{\mu}(\tau_t) + \mathbf{B}_0(\tau_t) \boldsymbol{\epsilon}_t + \mathbf{B}_1(\tau_t) \boldsymbol{\epsilon}_{t-1} + \mathbf{B}_2(\tau_t) \boldsymbol{\epsilon}_{t-2} + \dots$$

such that \mathbf{x}_t of (2.1) satisfies $\max_{t \geq 1} \{E \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|^\delta\}^{1/\delta} = O(T^{-1})$, where

$$\begin{aligned} \boldsymbol{\mu}(\tau) &= \mathbf{a}(\tau) + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j(\tau) \mathbf{a}(\tau), \quad \boldsymbol{\Psi}_j(\tau) = \mathbf{J} \boldsymbol{\Phi}^j(\tau) \mathbf{J}^\top \text{ for } j \geq 1, \\ \mathbf{J} &= [\mathbf{I}_d, \mathbf{0}_{d \times d(p-1)}], \quad \mathbf{B}_0(\tau) = \boldsymbol{\omega}(\tau), \quad \mathbf{B}_j(\tau) = \boldsymbol{\Psi}_j(\tau) \boldsymbol{\omega}(\tau), \\ \boldsymbol{\Phi}(\tau) &= \begin{bmatrix} \mathbf{A}_1(\tau) & \cdots & \mathbf{A}_{p-1}(\tau) & \mathbf{A}_p(\tau) \\ \mathbf{I}_d & \cdots & \mathbf{0}_d & \mathbf{0}_d \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_d & \cdots & \mathbf{I}_d & \mathbf{0}_d \end{bmatrix}. \end{aligned} \quad (2.3)$$

From Proposition 2.1, we can see that the $d \times 1$ vector of the orthogonalized impulse response function of a unit shock at time t to the j^{th} equation on \mathbf{x}_{t+n} is given by $\mathbf{B}_n(\tau_{t+n}) \mathbf{e}_j$, where \mathbf{e}_j is a

$d \times 1$ selection vector with unity as its j^{th} element and zeros elsewhere. Hence, the impulse responses produced by our model are deterministic functions of rescaled time, so that the TV-VAR model captures potential drifts in the transmission mechanism and produces impulse responses which are not history- and shock-dependent.

2.1 Estimation

To estimate $\{\mathbf{B}_j(\tau) : j \geq 0\}$, we need a joint central limit theorem for the estimators of the coefficients and the innovation covariance matrix. That said, we consider the estimation of $\mathbf{A}(\cdot)$ and $\mathbf{\Omega}(\cdot)$ using the local linear kernel method¹. Intuitively, when τ_t is in a small neighbourhood of τ , we can write (2.1) as

$$\mathbf{x}_t = \mathbf{A}(\tau_t)\mathbf{z}_{t-1} + \boldsymbol{\eta}_t \approx [\mathbf{A}(\tau), h\mathbf{A}^{(1)}(\tau)] \mathbf{z}_{t-1}^* + \boldsymbol{\eta}_t, \quad (2.4)$$

where $\mathbf{z}_{t-1} = [1, \mathbf{x}_{t-1}^\top, \dots, \mathbf{x}_{t-p}^\top]^\top$ and $\mathbf{z}_{t-1}^* = [\mathbf{z}_{t-1}^\top, \frac{\tau_t - \tau}{h} \mathbf{z}_{t-1}^\top]^\top$. The local linear estimators² of $\mathbf{A}(\tau)$ and $\mathbf{\Omega}(\tau)$ are then respectively given by

$$\begin{aligned} \text{vec}[\widehat{\mathbf{A}}(\tau)] &= [\mathbf{I}_{d^2 p+d}, \mathbf{0}_{d^2 p+d}] \cdot \left(\sum_{t=1}^T \mathbf{Z}_{t-1}^* \mathbf{Z}_{t-1}^{*\top} K_h(\tau_t - \tau) \right)^{-1} \sum_{t=1}^T \mathbf{Z}_{t-1}^* \mathbf{x}_t K_h(\tau_t - \tau), \\ \widehat{\mathbf{\Omega}}(\tau) &= \frac{1}{T} \sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top \omega_t(\tau), \end{aligned} \quad (2.5)$$

where $\mathbf{Z}_t^* = \mathbf{z}_t^* \otimes \mathbf{I}_d$, $\widehat{\boldsymbol{\eta}}_t = \mathbf{x}_t - \widehat{\mathbf{A}}(\tau_t)\mathbf{z}_{t-1}$, $\omega_t(\tau) = K_h(\tau_t - \tau) \frac{P_{h,2}(\tau) - \frac{\tau_t - \tau}{h} P_{h,1}(\tau)}{P_{h,0}(\tau)P_{h,2}(\tau) - P_{h,1}^2(\tau)}$ is the local linear weight, and $P_{h,k}(\tau) = \frac{1}{T} \sum_{t=1}^T \left(\frac{\tau_t - \tau}{h} \right)^k K_h(\tau_t - \tau)$ for $k = 0, 1, 2$.

We require the following conditions to hold for the kernel function and the bandwidth.

Assumption 3. *Let $K(\cdot)$ be a symmetric and positive kernel function defined on $[-1, 1]$ with $\int_{-1}^1 K(u)du = 1$. Moreover, $K(\cdot)$ is Lipschitz continuous on $[-1, 1]$. As $(T, h) \rightarrow (\infty, 0)$, $Th \rightarrow \infty$.*

With Assumption 3 in hand, we summarize the first theorem of this paper below.

Theorem 2.1. *Let Assumptions 1-3 hold. Suppose that $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s., and $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \rightarrow \infty$. Then*

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

In addition, suppose that conditional on \mathcal{F}_{t-1} , the third and fourth moments of $\boldsymbol{\epsilon}_t$ are identical to the corresponding unconditional moments a.s., and $Th^5 \rightarrow \alpha \in [0, \infty)$. Then the following two results also hold for $\forall \tau \in (0, 1)$:

¹The local linear kernel method allows us to address the so-called boundary effects of the kernel estimation when establishing the first result of Theorem 2.1 below. Alternatively, one may consider some boundary adjustment approaches as mentioned in Hong and Li (2005) and Connor et al. (2012).

²It is worth pointing out that the estimation of the covariance matrix using the local linear kernel method (such as the second estimator of (2.5)) is a non-trivial problem, and even has its own literature. We refer interested readers to Zhang and Wu (2012) for more details.

$$2. \sqrt{Th}\widehat{\mathbf{V}}^{-1/2}(\tau) \begin{bmatrix} \text{vec} \left(\widehat{\mathbf{A}}(\tau) - \mathbf{A}(\tau) - \frac{1}{2}h^2\tilde{c}_2\mathbf{A}^{(2)}(\tau) \right) \\ \text{vech} \left(\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau) - \frac{1}{2}h^2\tilde{c}_2\mathbf{\Omega}^{(2)}(\tau) \right) \end{bmatrix} \rightarrow_D N(\mathbf{0}, \mathbf{I}),$$

where $\mathbf{V}(\tau)$ and $\widehat{\mathbf{V}}(\tau)$ are defined in (A.3) and (A.5) for the sake of presentation.

The first result of Theorem 2.1 establishes a uniform convergence rate for $\widehat{\mathbf{A}}(\tau)$, which further allows us to establish a joint asymptotic distribution in the second result. If $\delta > 5$, the usual optimal bandwidth $h_{opt} = O(T^{-1/5})$ satisfies the condition $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \rightarrow \infty$.

2.2 On Impulse Responses

Having established the joint CLT in Theorem 2.1, we are now ready to study the impulse responses. As $\mathbf{\Omega}(\cdot) = \boldsymbol{\omega}(\cdot)\boldsymbol{\omega}^\top(\cdot)$, we cannot infer the elements of $\boldsymbol{\omega}(\cdot)$ unless certain identification restrictions are imposed. In the following, we consider two types of identification conditions: (i) the short-run timing restrictions, and (ii) the long-run restrictions. The economic interpretations of the two types of identification conditions can be found in Kilian and Lütkepohl (2017), and we do not repeat them here for the sake of space.

Under the short-run timing restrictions, $\boldsymbol{\omega}(\cdot)$ is a lower-triangular matrix. Thus, $\widehat{\boldsymbol{\omega}}(\tau)$ is chosen as the lower triangular matrix from the Cholesky decomposition of $\widehat{\mathbf{\Omega}}(\tau)$, i.e., $\widehat{\mathbf{\Omega}}(\tau) = \widehat{\boldsymbol{\omega}}(\tau)\widehat{\boldsymbol{\omega}}^\top(\tau)$. Alternatively, one can impose the conditions on the long-run impacts of the shocks (i.e., $\mathbf{B}(\tau)$ defined below). Specifically, define

$$\begin{aligned} \mathbf{B}(\tau) &:= \sum_{j=0}^{\infty} \mathbf{B}_j(\tau) = \left(\mathbf{I}_d - \sum_{i=1}^p \mathbf{A}_i(\tau) \right)^{-1} \boldsymbol{\omega}(\tau), \\ \boldsymbol{\Psi}(\tau) &:= \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j(\tau) = \left(\mathbf{I}_d - \sum_{i=1}^p \mathbf{A}_i(\tau) \right)^{-1}, \end{aligned} \quad (2.6)$$

where the last equalities of both lines follow in an obvious matter. Thus, the elements of $\mathbf{B}(\tau)$ may be recovered from $\mathbf{B}(\tau)\mathbf{B}^\top(\tau) = \boldsymbol{\Psi}(\tau)\mathbf{\Omega}(\tau)\boldsymbol{\Psi}^\top(\tau)$. It is then convenient to assume that $\mathbf{B}(\tau)$ is a lower-triangular matrix, so $\widehat{\mathbf{B}}(\tau)$ is chosen as the lower triangular matrix from the Cholesky decomposition of $\widehat{\boldsymbol{\Psi}}(\tau)\widehat{\mathbf{\Omega}}(\tau)\widehat{\boldsymbol{\Psi}}^\top(\tau)$, where $\widehat{\boldsymbol{\Psi}}(\tau)$ is defined in the same way as $\boldsymbol{\Psi}(\tau)$ in (2.3) but replacing $\mathbf{A}_i(\tau)$ with $\widehat{\mathbf{A}}_i(\tau)$. Under the long-run restrictions, $\widehat{\boldsymbol{\omega}}(\tau) = \widehat{\boldsymbol{\Psi}}^{-1}(\tau)\widehat{\mathbf{B}}(\tau)$.

Either way, the estimator of the impulse response function $\mathbf{B}_j(\tau)$ for each given $j \geq 0$ is given by

$$\widehat{\mathbf{B}}_j(\tau) = \widehat{\boldsymbol{\Psi}}_j(\tau)\widehat{\boldsymbol{\omega}}(\tau), \quad (2.7)$$

where $\widehat{\boldsymbol{\Psi}}_j(\tau) = \mathbf{J}\widehat{\boldsymbol{\Phi}}^j(\tau)\mathbf{J}^\top$, \mathbf{J} is defined in Proposition 2.1, and $\widehat{\boldsymbol{\Phi}}^j(\tau)$ is defined under (A.6).

We summarize the asymptotic properties of the estimation of the impulse responses in the following theorem.

Theorem 2.2. *Under the conditions of Theorem 2.1. For any fixed integer $j \geq 0$*

$$\sqrt{Th} \left(\text{vec} \left(\widehat{\mathbf{B}}_j(\tau) - \mathbf{B}_j(\tau) \right) - \frac{1}{2} h^2 \tilde{c}_2 \mathbf{B}_j^{(2)}(\tau) \right) \rightarrow_D N(0, \boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)),$$

where

$$\begin{aligned} \mathbf{B}_j^{(2)}(\tau) &= \mathbf{C}_{j,1}(\tau) \text{vec}(\mathbf{A}^{(2)}(\tau)) + \mathbf{C}_{j,2}(\tau) \text{vech}(\boldsymbol{\Omega}^{(2)}(\tau)), \\ \boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau) &= [\mathbf{C}_{j,1}(\tau), \mathbf{C}_{j,2}(\tau)] \mathbf{V}(\tau) [\mathbf{C}_{j,1}(\tau), \mathbf{C}_{j,2}(\tau)]^\top. \end{aligned}$$

Specifically, we have the following expressions.

1. Under the short-run timing restrictions, we have

$$\begin{aligned} \mathbf{C}_{0,1}(\tau) &= 0, \\ \mathbf{C}_{j,1}(\tau) &= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_d) \left(\sum_{m=0}^{j-1} \mathbf{J}(\boldsymbol{\Phi}^\top(\tau))^{j-1-m} \otimes \boldsymbol{\Psi}_m(\tau) \right) [\mathbf{0}_{d^2 p \times d}, \mathbf{I}_{d^2 p}], \quad j \geq 1, \\ \mathbf{C}_{j,2}(\tau) &= (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \mathbf{L}_d^\top (\mathbf{L}_d \mathbf{N}_1(\tau) \mathbf{L}_d^\top)^{-1}, \quad j \geq 0, \end{aligned}$$

in which $\mathbf{N}_1(\tau) = (\mathbf{I}_{d^2} + \mathbf{K}_{d,d})(\boldsymbol{\omega}(\tau) \otimes \mathbf{I}_d)$, the elimination matrix \mathbf{L}_d satisfies that $\text{vech}(\mathbf{F}) = \mathbf{L}_d \text{vec}(\mathbf{F})$ for any $d \times d$ matrix \mathbf{F} , and the commutation matrix $\mathbf{K}_{m,n}$ satisfies $\mathbf{K}_{m,n} \text{vec}(\mathbf{G}) = \text{vec}(\mathbf{G}^\top)$ for any $m \times n$ matrix \mathbf{G} .

2. Under the long-run restrictions, we have

$$\begin{aligned} \mathbf{C}_{0,1}(\tau) &= (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) (\mathbf{N}_1^\top(\tau) \mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau) \mathbf{N}_2(\tau))^{-1} \mathbf{N}_2^\top(\tau) \mathbf{D}_2(\tau), \\ \mathbf{C}_{j,1}(\tau) &= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_d) \left(\sum_{m=0}^{j-1} \mathbf{J}(\boldsymbol{\Phi}^\top(\tau))^{j-1-m} \otimes \boldsymbol{\Psi}_m(\tau) \right) [\mathbf{0}_{d^2 p \times d}, \mathbf{I}_{d^2 p}] + (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \\ &\quad \times (\mathbf{N}_1^\top(\tau) \mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau) \mathbf{N}_2(\tau))^{-1} \mathbf{N}_2^\top(\tau) \mathbf{D}_2(\tau) [\mathbf{0}_{d^2 p \times d}, \mathbf{I}_{d^2 p}], \quad j \geq 1, \\ \mathbf{C}_{j,2}(\tau) &= (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) (\mathbf{N}_1^\top(\tau) \mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau) \mathbf{N}_2(\tau))^{-1} \mathbf{N}_1^\top(\tau) \mathbf{D}_1, \quad j \geq 0, \end{aligned}$$

in which $\mathbf{N}_2(\tau) = \mathbf{Q} (\mathbf{I}_d \otimes \mathbf{A}_\tau^{-1}(1))$, $\mathbf{D}_2(\tau) = \mathbf{Q} (\mathbf{B}^\top(\tau) \otimes \mathbf{A}_\tau^{-1}(1)) \nabla_{\boldsymbol{\alpha}(\tau)} \mathbf{A}_\tau(1)$ with $\mathbf{A}_\tau(1) = \mathbf{I}_d - \sum_{i=1}^p \mathbf{A}_i(\tau)$ and $\nabla_{\boldsymbol{\alpha}(\tau)} \mathbf{A}_\tau(1) = -[\mathbf{I}_{d^2}, \dots, \mathbf{I}_{d^2}] (d^2 \times d^2 p)$, the duplication matrix \mathbf{D}_1 satisfies $\text{vec}(\boldsymbol{\Omega}(\tau)) = \mathbf{D}_1 \text{vech}(\boldsymbol{\Omega}(\tau))$, and \mathbf{Q} is a $d(d-1)/2 \times d^2$ selection matrix of 0 and 1 such that $\mathbf{Q} \text{vec}(\mathbf{B}(\tau)) = 0$.

For ease of presentation, we provide the definitions of the respective estimators of $\mathbf{V}(\tau)$ and $\boldsymbol{\Phi}(\tau)$ (i.e., $\widehat{\mathbf{V}}(\tau)$ and $\widehat{\boldsymbol{\Phi}}(\tau)$) in Appendix A.1. It is easy to see that $\widehat{\boldsymbol{\Phi}}(\tau) \rightarrow_P \boldsymbol{\Phi}(\tau)$, $\widehat{\boldsymbol{\omega}}(\tau) \rightarrow_P \boldsymbol{\omega}(\tau)$, and $\widehat{\mathbf{V}}(\tau) \rightarrow_P \mathbf{V}(\tau)$ by Theorem 2.1. As a result, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{B}_j}(\tau) \rightarrow_P \boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)$, where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{B}_j}(\tau)$ has a form identical to $\boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)$ but replacing $\boldsymbol{\Phi}(\tau)$, $\boldsymbol{\omega}(\tau)$ and $\mathbf{V}(\tau)$ with their estimators, respectively.

2.3 Selection of the Optimal Lag

We next consider the choice of the optimal number of lags, i.e., the estimation of p . Specifically, we consider the minimization of the next information criterion.

$$\hat{\mathbf{p}} = \underset{1 \leq p \leq P}{\operatorname{argmin}} (\log \{\operatorname{RSS}(\mathbf{p})\} + \mathbf{p} \cdot \chi_T) \equiv \underset{1 \leq p \leq P}{\operatorname{argmin}} \operatorname{IC}(\mathbf{p}), \quad (2.8)$$

where $\operatorname{RSS}(\mathbf{p}) = \frac{1}{T} \sum_{t=1}^T \hat{\boldsymbol{\eta}}_{\mathbf{p},t}^\top \hat{\boldsymbol{\eta}}_{\mathbf{p},t}$, χ_T is the penalty term, $\hat{\boldsymbol{\eta}}_{\mathbf{p},t}$ is the value of $\hat{\boldsymbol{\eta}}_t$ by letting the lag be \mathbf{p} , and P is a sufficiently large fixed positive integer. The next theorem shows the validity of (2.8).

Theorem 2.3. *Let Assumptions 1-3 hold. Suppose $\frac{T^{1-\frac{4}{3}}h}{\log T} \rightarrow \infty$, $\max_{t \geq 1} E(\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}) < \infty$ a.s., $\chi_T \rightarrow 0$, and $c_T^{-2} \chi_T \rightarrow \infty$, where $c_T = h^2 + (\frac{\log T}{Th})^{1/2}$. Then $\Pr(\hat{\mathbf{p}} = p) \rightarrow 1$.*

In view of the conditions on χ_T , a natural choice is

$$\chi_T = \max \left\{ h^4, \frac{\log T}{Th} \right\} \cdot \log(\log(Th)).$$

To this end, we have established the asymptotic properties of the proposed estimators. In the following, we discuss some model specification issues when building (2.1).

3 Testing for Parameter Stability

To close our investigation, we consider testing whether some (if not all) components of the coefficient matrices are time-invariant, which as explained in the introduction may help settle the discrepancy on the ongoing debate about the high inflation in the U.S. during 1970-1980.

For model (2.1), we are specifically interested in testing

$$\mathbb{H}_0 : \mathbf{C}\boldsymbol{\beta}(\cdot) = \mathbf{c} \text{ for some unknown } \mathbf{c} \in \mathbb{R}^s, \quad (3.1)$$

where $\boldsymbol{\beta}(\tau) := \operatorname{vec}(\mathbf{A}(\tau))$ and \mathbf{C} is a selection matrix. Practically, the choice of \mathbf{C} should be theory/application driven, and \mathbf{c} needs to be estimated. For example, in the context of monetary policy analysis (Primiceri, 2005), one can let $\mathbf{C} = [\mathbf{0}_{d^2 p \times d}, \mathbf{I}_{d^2 p}]$ to test whether the policy transmission mechanism is varying over time.

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

$$\hat{Q}_{\mathbf{C}, \mathbf{H}} = \int_0^1 \left\{ \mathbf{C}\hat{\boldsymbol{\beta}}(\tau) - \hat{\mathbf{c}} \right\}^\top \mathbf{H}(\tau) \left\{ \mathbf{C}\hat{\boldsymbol{\beta}}(\tau) - \hat{\mathbf{c}} \right\} d\tau, \quad (3.2)$$

where $\hat{\boldsymbol{\beta}}(\cdot) := \operatorname{vec}[\hat{\mathbf{A}}(\cdot)]$ should be obvious, and $\hat{\mathbf{c}}$ is the estimator of \mathbf{c} to be constructed below. In (3.2), $\mathbf{H}(\cdot)$ is an $s \times s$ positive definite weighting matrix, and is typically set as the precision matrix associated with $\hat{\boldsymbol{\beta}}(\cdot)$.

Let \mathbf{C} be a pre-specified matrix and $\mathbf{c} = \int_0^1 \mathbf{C}\boldsymbol{\beta}(\tau)d\tau$. Then a natural estimator of \mathbf{c} is given by

$$\widehat{\mathbf{c}} = \int_0^1 \mathbf{C}\widehat{\boldsymbol{\beta}}(\tau)d\tau, \quad (3.3)$$

of which the asymptotic properties are summarized in the following lemma.

Lemma 3.1. *Let Assumptions 1-3 hold. Suppose further that $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s., $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \rightarrow \infty$, each element of $\mathbf{A}(\cdot)$ has finite third-order derivative, $Th^2/(\log T)^2 \rightarrow \infty$, and $Th^6 \rightarrow 0$. As $T \rightarrow \infty$, we have*

$$\sqrt{T} \left(\widehat{\mathbf{c}} - \mathbf{c} - \frac{1}{2}h^2\tilde{c}_2 \int_0^1 \mathbf{C}\boldsymbol{\beta}^{(2)}(\tau)d\tau \right) \rightarrow_D N \left(\mathbf{0}, \int_0^1 \mathbf{C}\mathbf{V}_{\boldsymbol{\beta}}(\tau)\mathbf{C}^\top d\tau \right),$$

where $\mathbf{V}_{\boldsymbol{\beta}}(\tau) := \boldsymbol{\Sigma}^{-1}(\tau) \otimes \boldsymbol{\Omega}(\tau)$ and $\boldsymbol{\Sigma}(\tau)$ is defined in (A.1).

Lemma 3.1 indicates that under the null, $\widehat{\mathbf{c}}$ will achieve a parametric rate (i.e., \sqrt{T}). This is important for the finite sample study, since more reliable information can be retrieved from the estimators.

Having established Lemma 3.1, we move on to investigate the test statistic (3.2), and examine the local alternatives. After some tedious development, one actually can show that

$$\widehat{Q}_{\mathbf{C},\mathbf{H}} = \frac{1}{Th} \cdot \frac{\tilde{v}_0}{T} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \mathbf{H}_0(\tau_t) \mathbf{Z}_{t-1} \boldsymbol{\eta}_t + O_P \left(\frac{1}{T\sqrt{h}} \right) \quad (3.4)$$

where $\mathbf{Z}_t = \mathbf{z}_t \otimes \mathbf{I}_d$, $\mathbf{H}_0(\tau) = \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau)\mathbf{C}^\top \mathbf{H}(\tau)\mathbf{C}\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau)$, $\boldsymbol{\Sigma}_{\mathbf{Z}}(\tau) = \boldsymbol{\Sigma}(\tau) \otimes \mathbf{I}_d$. Clearly,

$$\frac{\tilde{v}_0}{T} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \mathbf{H}_0(\tau_t) \mathbf{Z}_{t-1} \boldsymbol{\eta}_t \quad (3.5)$$

converges to a fixed value in probability, so the first term on the right hand side of (3.4) is the leading one. Moreover, $\frac{\tilde{v}_0}{T} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \mathbf{H}_0(\tau_t) \mathbf{Z}_{t-1} \boldsymbol{\eta}_t$ can be considered as a weighted average of the second moment of $\boldsymbol{\eta}_t$, it thus 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. The term which really yields an asymptotic distribution is in fact the second part on the right hand side of (3.4). Thus, we present the asymptotic distribution of the test statistic as follows.

Theorem 3.1. *Let the conditions of Lemma 3.1 hold. Under the null hypothesis, if $Th^{5.5} \rightarrow 0$ and $E[\|\boldsymbol{\epsilon}_t\|^\delta | \mathcal{F}_{t-1}] < \infty$ a.s., we have*

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

where $\widehat{\mathbf{H}}(\tau) = (\mathbf{C}\widehat{\mathbf{V}}_{\boldsymbol{\beta}}(\tau)\mathbf{C}^\top)^{-1}$, $\widehat{\mathbf{V}}_{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\Sigma}}^{-1}(\tau) \otimes \widehat{\boldsymbol{\Omega}}(\tau)$, $C_B = \int_0^2 \left(\int_{-1}^{1-v} K(u)K(u+v)du \right)^2 dv$, and $\widehat{\boldsymbol{\Sigma}}(\tau)$ is defined in (A.5).

Theorem 3.1 states that the test statistic converges to a normal distribution. The term $s\tilde{v}_0$ is the limit associated with (3.5). Due to the use of $\widehat{\mathbf{H}}(\cdot)$, the second moment of $\boldsymbol{\eta}_t$ 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 (3.2), any departure from the true value will eventually yield a squared term when analysing the asymptotic power. Therefore, the null of (3.1) will be rejected at the level α if

$$\widehat{Q}_{\mathbf{C}, \widehat{\mathbf{H}}}^* = \frac{T\sqrt{h} \left(\widehat{Q}_{\mathbf{C}, \widehat{\mathbf{H}}} - \frac{1}{Th} s\tilde{v}_0 \right)}{\sqrt{4sC_B}} > q_{1-\alpha}, \quad (3.6)$$

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

In what follows, we consider a sequence of local alternatives of the form:

$$\mathbb{H}_1 : \mathbf{C}\boldsymbol{\beta}(\tau) = \mathbf{c} + d_T \mathbf{f}(\tau), \quad (3.7)$$

where $\mathbf{f}(\tau)$ is a twice continuously differentiable vector of functions, and $d_T \rightarrow 0$. The term $d_T \mathbf{f}(\tau)$ characterizes the departure of the time-varying coefficient $\mathbf{C}\boldsymbol{\beta}(\tau)$ from the constant \mathbf{c} . Following the development of Theorem 3.1, it is straightforward to obtain the following corollary.

Corollary 3.1. *Let the conditions of Theorem 3.1 hold. Under the \mathbb{H}_1 of (3.7), if $d_T = T^{-1/2}h^{-1/4}$, then*

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

where $\delta_1 = \int_0^1 \mathbf{f}(\tau)^\top (\mathbf{C}\mathbf{V}_\beta(\tau)\mathbf{C}^\top)^{-1} \mathbf{f}(\tau) d\tau$.

Moreover, we have $\Pr \left(\widehat{Q}_{\mathbf{C}, \widehat{\mathbf{H}}}^* > q_{1-\alpha} \right) \rightarrow \Phi \left(q_\alpha + \frac{\delta_1}{2\sqrt{sC_B}} \right)$.

Corollary 3.1 shows that the test has a non-trivial power against \mathbb{H}_1 when $d_T = T^{-1/2}h^{-1/4}$. If $T^{-1/2}h^{-1/4} = o(d_T)$, the power of the test converges to 1, i.e.,

$$\Pr(\widehat{Q}_{\mathbf{C}, \widehat{\mathbf{H}}}^* > q_{1-\alpha}) \rightarrow 1.$$

Before we conclude this section, we would like to add some comments on the assumptions imposed and the main results established in relation to the relevant literature. The construction of the proposed test is similar to those discussed in Zhang and Wu (2012) and then Truquet (2017). Because we have developed and then employed Proposition 2.1 for the time-varying VMA(∞) representation, the assumptions, such as requiring $Th^{5.5} \rightarrow 0$, are less restrictive than those assumed in the relevant literature, see, for example, $Th^{3.5} \rightarrow 0$ by Truquet (2017). As a consequence, the main techniques employed in our proofs may be of general interest and applicability in dealing with similar problems.

In what follows, we will examine the finite sample performance of the asymptotic properties for the proposed estimators and test statistic by simulation studies.

4 Simulation

In this section, we first provide some details of the numerical implementation in Section 4.1, and then respectively examine the estimation and hypothesis testing in Sections 4.2 and 4.3.

4.1 Numerical Implementation

Throughout the numerical studies, Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ is adopted.

When selecting the optimal lag by (2.8), the bandwidth \hat{h}_{cv} is always chosen by minimizing the following cross-validation criterion function for each \mathbf{p} .

$$\hat{h}_{cv} = \arg \min_h \sum_{t=1}^T \left\| \mathbf{x}_t - \hat{\mathbf{a}}_{-t}(\tau_t) - \sum_{j=1}^{\mathbf{p}} \hat{\mathbf{A}}_{j,-t}(\tau_t) \mathbf{x}_{t-j} \right\|^2, \quad (4.1)$$

where $\hat{\mathbf{a}}_{-t}(\cdot)$, and $\hat{\mathbf{A}}_{j,-t}(\cdot)$ are obtained using (2.5) but leaving the t^{th} observation out. Once $\hat{\mathbf{p}}$ and \hat{h}_{cv} are obtained, the estimation procedure is relatively straightforward. It is pointed out that it might be possible to extend the Bayesian bandwidth selection method proposed for the conventional time-varying regression model by Cheng et al. (2019) to the time-varying VAR setting under study for an optimal choice of the bandwidth parameter.

We then comment on the testing procedure. To improve the finite sample performance of the test, we propose a simulation-assisted testing procedure. A similar procedure has also been adopted by Zhang and Wu (2012) and Truquet (2017) to for the same purpose in the context of univariate time-varying models. Alternatively, one may consider the moving block bootstrap as in Jentsch and Lunsford (2021). For simplicity, we adopt the former as follows.

Algorithm — a simulation-assisted testing procedure

Step 1: Use the sample $\{\mathbf{x}_t\}$ to estimate the unrestricted model and the restricted model, and then compute $\hat{Q}_{\mathbf{C},\hat{\mathbf{H}}}$ based on (3.2).

Step 2: Generate i.i.d. standard multivariate normal random vectors $\{\mathbf{x}_t^*\}$.

Step 3: Compute the bootstrap statistic $\tilde{Q}_{\mathbf{C},\hat{\mathbf{H}}}^b$ in the same way as $\hat{Q}_{\mathbf{C},\hat{\mathbf{H}}}$, with $\{\mathbf{x}_t^*\}$ replacing the original sample $\{\mathbf{x}_t\}$.

Step 4: Repeat Steps 2-3 B times to obtain B bootstrap test statistics $\{\tilde{Q}_{\mathbf{C},\hat{\mathbf{H}}}^b\}_{b=1}^B$, as well as its empirical quantile $\hat{q}_{1-\alpha}$. We reject the null hypothesis (3.1) at level α if $\hat{Q}_{\mathbf{C},\hat{\mathbf{H}}} > \hat{q}_{1-\alpha}$.

4.2 Examining the Model Estimation

We now examine the finite sample performance of the theoretical findings. The data generating process (DGP) is as follows.

$$\mathbf{x}_t = \mathbf{a}(\tau_t) + \mathbf{A}_1(\tau_t)\mathbf{x}_{t-1} + \mathbf{A}_2(\tau_t)\mathbf{x}_{t-2} + \boldsymbol{\eta}_t \quad \text{with} \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t \quad \text{for} \quad t = 1, \dots, T, \quad (4.2)$$

where ϵ_t 's are i.i.d. draws from $N(\mathbf{0}_{2 \times 1}, \mathbf{I}_2)$,

$$\begin{aligned} \mathbf{a}(\tau) &= [0.5 \sin(2\pi\tau), 0.5 \cos(2\pi\tau)]^\top, \\ \mathbf{A}_1(\tau) &= \begin{bmatrix} 0.8 \exp(-0.5 + \tau) & 0.8(\tau - 0.5)^3 \\ 0.8(\tau - 0.5)^3 & 0.8 + 0.3 \sin(\pi\tau) \end{bmatrix}, \\ \mathbf{A}_2(\tau) &= \begin{bmatrix} -0.2 \exp(-0.5 + \tau) & 0.8(\tau - 0.5)^2 \\ 0.8(\tau - 0.5)^2 & -0.4 + 0.3 \cos(\pi\tau) \end{bmatrix}, \\ \boldsymbol{\omega}(\tau) &= \begin{bmatrix} 1.5 + 0.2 \exp(0.5 - \tau) & 0 \\ 0.1 \exp(0.5 - \tau) & 1.5 + 0.5(\tau - 0.5)^2 \end{bmatrix}. \end{aligned}$$

Let the sample size be $T \in \{200, 400, 800\}$, and conduct 1000 replications for each choice of T .

First, we check whether the coefficient matrices $\mathbf{A}_1(\tau)$ and $\mathbf{A}_2(\tau)$ satisfy Assumption 1.1. Thus, for each generated dataset, we compute the largest eigenvalue of the true companion matrix $\boldsymbol{\Phi}(\tau)$ in $\tau \in [0, 1]$, which varies from 0.54 to 0.88 over replications indicating the validity of Assumption 1.1.

Next, we report the percentages of $\hat{\mathbf{p}} < 2$, $\hat{\mathbf{p}} = 2$, and $\hat{\mathbf{p}} > 2$ respectively based on 1000 replications. Table 1 shows that the information criterion (2.8) performs reasonably well, as the percentages associated with $\hat{\mathbf{p}} = 2$ are sufficiently close to 1 even for $T = 200$.

Table 1: The percentages of $\hat{\mathbf{p}} < 2$, $\hat{\mathbf{p}} = 2$, and $\hat{\mathbf{p}} > 2$

T	$\hat{\mathbf{p}} < 2$	$\hat{\mathbf{p}} = 2$	$\hat{\mathbf{p}} > 2$
200	0.004	0.976	0.020
400	0.004	0.986	0.010
800	0.000	1.000	0.000

Finally, we evaluate the estimates of $\mathbf{A}(\tau)$, $\boldsymbol{\Omega}(\tau)$, as well as the estimates of the impulse responses (say, $\mathbf{B}_1(\tau)$ and $\mathbf{B}_5(\tau)$ without loss of generality) based on the short-run timing restrictions. For each parameter of interest, we calculate the root mean square error (RMSE) as follows:

$$\left\{ \frac{1}{1000T} \sum_{n=1}^{1000} \sum_{t=1}^T \|\hat{\boldsymbol{\theta}}^{(n)}(\tau_t) - \boldsymbol{\theta}(\tau_t)\|^2 \right\}^{1/2},$$

where $\boldsymbol{\theta}(\cdot) \in \{\mathbf{A}(\cdot), \boldsymbol{\Omega}(\cdot), \mathbf{B}_1(\tau), \mathbf{B}_5(\tau)\}$, and $\hat{\boldsymbol{\theta}}^{(n)}(\tau)$ is the estimate of $\boldsymbol{\theta}(\tau)$ for the n^{th} replication. Of interest, we also report the finite sample coverage probabilities of the confidence intervals. The nominal coverage is 95%. Given $\boldsymbol{\theta}(\cdot)$, for each generated dataset, the coverage probability is first calculated for each functional component of $\boldsymbol{\theta}(\cdot)$ over the grid points $\{\tau_t, t = 1, \dots, T\}$, and then we further take an average across the elements of $\boldsymbol{\theta}(\cdot)$. After 1000 replications, we present the averaged value of these coverage probabilities in Table 2.

As shown in Table 2, the RMSEs decrease as the sample size increases. The finite sample

coverage probabilities are smaller than their nominal level when $T = 200$, but are fairly close to 95% as $T = 800$.

Table 2: The RMSEs and the empirical coverage probabilities (in parentheses)

T	$\mathbf{A}(\tau)$	$\mathbf{\Omega}(\tau)$	$\mathbf{B}_1(\tau)$	$\mathbf{B}_5(\tau)$
200	0.54 (0.89)	0.83 (0.87)	0.46 (0.87)	0.31 (0.89)
400	0.40 (0.91)	0.71 (0.91)	0.30 (0.91)	0.34 (0.89)
800	0.29 (0.92)	0.62 (0.93)	0.29 (0.92)	0.30 (0.90)

4.3 Examining the Parameter Stability Testing

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

$$\mathbf{x}_t = \mathbf{A}_1(\tau_t)\mathbf{x}_{t-1} + \mathbf{A}_2(\tau_t)\mathbf{x}_{t-2} + \boldsymbol{\eta}_t, \quad (4.3)$$

where $\mathbf{A}_2(\cdot)$ and $\boldsymbol{\eta}_t$ are generated in the same way as in Section 4.2, and

$$\mathbf{A}_1(\tau) = \begin{bmatrix} 0.4 & -0.1 \\ -0.1 & 0.4 \end{bmatrix} + b \times d_T \times \begin{bmatrix} 2\exp(\tau - 1) - 1 & \exp(\tau - 1) - 1 \\ \exp(\tau - 1) - 1 & 2\exp(\tau - 1) - 1 \end{bmatrix}$$

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. We use the proposed testing procedure to test whether the coefficient $\mathbf{A}_1(\cdot)$ is time-varying.

Again, we let $T \in \{200, 400, 800\}$ and conduct 1000 replications for each choice of T . We use the simulation-assisted testing procedure to get the empirical critical value $\widehat{q}_{1-\alpha}$ after 1000 bootstrap replications. We consider a sequence of bandwidths to check the robustness of the proposed test with respect to the bandwidth choice:

$$h = \alpha_1 T^{-1/5}, \quad \alpha_1 = 0.2, 0.4, \dots, 1.8. \quad (4.4)$$

Table 3 reports the rejection rates at the 5% and 10% nominal levels. A few facts emerge from the table. First, our test has reasonable sizes using empirical critical values obtained by the bootstrap procedure. Second, the size behaviour of our test is not sensitive to the choices of bandwidths. As discussed in Gao and Gijbels (2008), the estimation-based optimal bandwidths may also be optimal for testing purposes, so for simplicity one can use the rule-of-thumb in practice. Third, the local power of our test increases rapidly as b increases.

5 Empirical Study

In this section, we discuss the transmission mechanism of the monetary policy. As well documented, inflation is higher and more volatile in the U.S. during 1970-1980, but substantially decreases in

Table 3: Size and power evaluation

		5%			10%			
	Bandwidth	T	200	400	800	200	400	800
$b = 0$ (size)	$0.2T^{-1/5}$		0.046	0.060	0.058	0.081	0.113	0.117
	$0.4T^{-1/5}$		0.044	0.050	0.066	0.095	0.093	0.104
	$0.6T^{-1/5}$		0.057	0.070	0.046	0.110	0.121	0.102
	$0.8T^{-1/5}$		0.056	0.066	0.050	0.113	0.131	0.105
	$1.0T^{-1/5}$		0.033	0.060	0.047	0.066	0.104	0.100
	$1.2T^{-1/5}$		0.055	0.042	0.042	0.115	0.093	0.086
	$1.4T^{-1/5}$		0.039	0.045	0.043	0.085	0.103	0.078
	$1.6T^{-1/5}$		0.065	0.047	0.038	0.114	0.120	0.088
	$1.8T^{-1/5}$		0.041	0.066	0.046	0.094	0.117	0.088
$b = 2$ (local power)	$0.2T^{-1/5}$		0.095	0.103	0.143	0.177	0.189	0.242
	$0.4T^{-1/5}$		0.145	0.155	0.162	0.204	0.247	0.236
	$0.6T^{-1/5}$		0.148	0.129	0.170	0.208	0.255	0.239
	$0.8T^{-1/5}$		0.124	0.168	0.168	0.233	0.258	0.240
	$1.0T^{-1/5}$		0.144	0.176	0.145	0.230	0.250	0.225
	$1.2T^{-1/5}$		0.123	0.189	0.158	0.196	0.274	0.246
	$1.4T^{-1/5}$		0.160	0.173	0.183	0.246	0.275	0.279
	$1.6T^{-1/5}$		0.142	0.139	0.189	0.208	0.218	0.292
	$1.8T^{-1/5}$		0.172	0.170	0.194	0.269	0.268	0.287
$b = 4$ (local power)	$0.2T^{-1/5}$		0.477	0.448	0.655	0.635	0.604	0.783
	$0.4T^{-1/5}$		0.340	0.561	0.648	0.513	0.705	0.758
	$0.6T^{-1/5}$		0.428	0.582	0.708	0.583	0.709	0.803
	$0.8T^{-1/5}$		0.506	0.579	0.640	0.615	0.701	0.773
	$1.0T^{-1/5}$		0.501	0.579	0.585	0.598	0.699	0.695
	$1.2T^{-1/5}$		0.473	0.527	0.620	0.592	0.657	0.728
	$1.4T^{-1/5}$		0.482	0.532	0.553	0.614	0.648	0.655
	$1.6T^{-1/5}$		0.519	0.559	0.583	0.617	0.664	0.665
	$1.8T^{-1/5}$		0.460	0.523	0.613	0.584	0.653	0.720

the subsequent period, which is often referred to as the Great Moderation (Primiceri, 2005). Two explanations (i.e., bad policy and bad luck) have been debated repeatedly in the literature. The first explanation focuses on the changes in the transmission mechanism (e.g., Cogley and Sargent, 2005), while the second regards it as a consequence of changes in the size of exogenous shocks (e.g., Sims and Zha, 2006). In what follows, we revisit the arguments associated with the Great Moderation using our approach. The estimation is conducted in exactly the same way as in Section 4, so we no longer repeat the details.

First, we estimate the time-varying VAR(p) model using three commonly adopted macroeconomic variables of the literature (Primiceri, 2005; Cogley et al., 2010), which are the inflation rate (measured by 100 times the year-over-year log change in the GDP deflator), the unemployment

rate representing the non-policy block, and the interest rate (measured by the average value for the Federal funds rates over the quarter) representing the monetary policy block. Following Primiceri (2005), the short-run timing restrictions are employed to identify the monetary policy shocks. The interest rate is included as the third variable and being treated as the monetary policy instrument. The identification requirement is that the monetary policy actions affect the inflation and the unemployment with at least one period of lag (i.e., short-run timing restrictions). The data are quarterly observations measured at an annual rate from 1954:Q3 to 2015:Q4, which are collected from the Federal Reserve Bank of St. Louis economic database. Figure 1 plots the three variables.

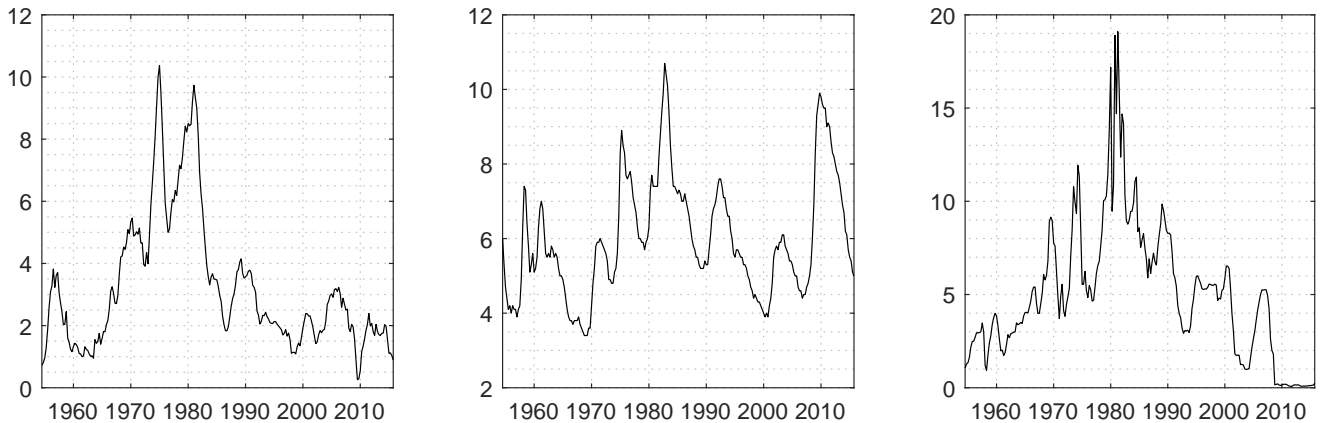


Figure 1: Plots of the inflation (left), the unemployment rate (middle) and the interest rate (right)

For the time-varying VAR(p) model, the optimal lag is $\hat{p} = 3$ by our approach, while it is often assumed to be known with the values varying from 2 to 4 in the literature. Thus, the following analyses focus on the time-varying VAR(3) model (referred to as TV-VAR(3) hereafter). We then conduct a robustness check to see whether the innovation process ϵ_t exhibits serial correlation. We use the Breusch-Godfrey LM test (cf., Breusch, 1978; Godfrey, 1978) for testing the first-order autocorrelation of ϵ_t 's. The null hypothesis is $H_0 : E(\epsilon_t \epsilon_{t-1}^\top) = 0$. Based on the estimates $\hat{\epsilon}_t = \hat{\omega}^{-1}(\tau_t) \hat{\eta}_t$, the p -value is 0.81 suggesting that the TV-VAR(3) model fits the data quite well.

Before investigating the time-varying effects of monetary policy on real economy, it is reasonable to check whether the VAR coefficients (i.e., the policy transmission mechanism) are time-varying. To this end, we employ the proposed test statistic to examine the constancy of model coefficients, and summarize the results in Table 4. Specifically, we first set \mathbf{C} to be an identity matrix for selecting between parametric and time-varying VAR models. From Table 4 (the row labelled ‘‘Constancy’’), we conclude that the VAR coefficients are not constant, implying that there exists significant time-variations in policy transmission mechanism. We then apply the testing procedures to examine whether the selected variables (i.e., intercept, \mathbf{x}_{t-1} , \mathbf{x}_{t-2} , \mathbf{x}_{t-3}) have time-varying contributions. By Table 4, at the 5% significance level we conclude that all $\mathbf{a}(\cdot)$, $\mathbf{A}_1(\cdot)$, $\mathbf{A}_2(\cdot)$ and $\mathbf{A}_3(\cdot)$ are time-varying by examining them individually³, suggesting that the TV-VAR(3) is more appropriate than a constant VAR model.

³Certainly, one may examine each element of these matrices. However, it will lead to a quite lengthy presentation. In order not to deviate from our main goal, we no longer conduct more testing along this line.

Table 4: Summary statistics of the test

	test statistic	p-value
Constancy	341.27	0.00
$\mathbf{a}(\cdot)$	50.15	0.00
$\mathbf{A}_1(\cdot)$	63.02	0.00
$\mathbf{A}_2(\cdot)$	23.27	0.00
$\mathbf{A}_3(\cdot)$	17.23	0.01

We now consider the changes in the size of exogenous shocks. Figure 2 plots the estimated residuals $\hat{\boldsymbol{\eta}}_t$. As can be seen, the magnitudes of estimated residuals change over time. For example, the magnitudes of the estimated residuals in the interest rate component increase from 1970 to 1980, but decrease quickly after 1980. The evidence favours the TV-VAR(3) model which accounts for heteroscedasticity. More precisely, Figure 3 plots the time-varying standard deviation of the estimated residuals $\hat{\boldsymbol{\eta}}_t$ as well as the associated 95% confidence intervals. We can see that a decline in unconditional volatilities of exogenous shocks after 1980 in every subplot of Figure 3. Thus, our results support the explanation of “bad luck” (e.g., Primiceri, 2005; Sims and Zha, 2006).

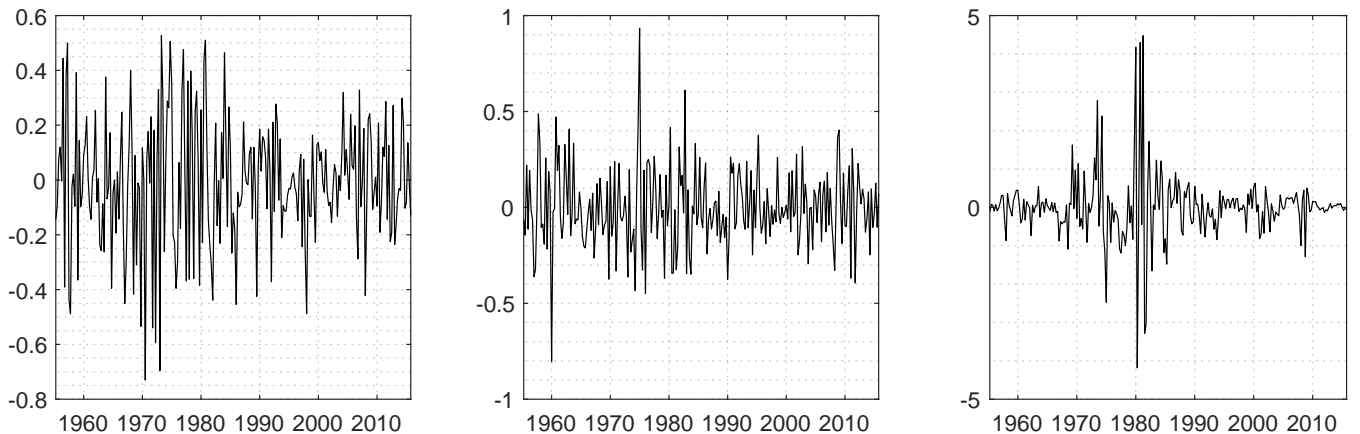


Figure 2: Estimation residuals in the inflation equation (left), the unemployment equation (middle) and the interest rate equation (right).

We then discuss how inflation and unemployment respond to monetary policy shocks. Figure 4 plots the time-varying cumulative impulse responses⁴ of inflation and unemployment to a monetary shock subject to the short-run timing restrictions. Clearly, these responses vary over time, indicating a substantial time-variation in the policy transmission mechanism. The result differs from Primiceri (2005) that has found no evidence of time variation in the response of the economy to monetary policy shocks using a Bayesian approach. Interestingly, our results show that the responses of inflation exhibit a price puzzle during 1970-1980, which suggests that inflation increases in response to a monetary tightening although standard macroeconomic theory predicts

⁴The cumulative impulse responses are constructed in exactly the same way as in Primiceri (2005), and the standard errors are computed based on Theorem 2.2.

the opposite. Therefore, our results also support the explanation of “bad policy” (e.g., Cogley and Sargent, 2005).

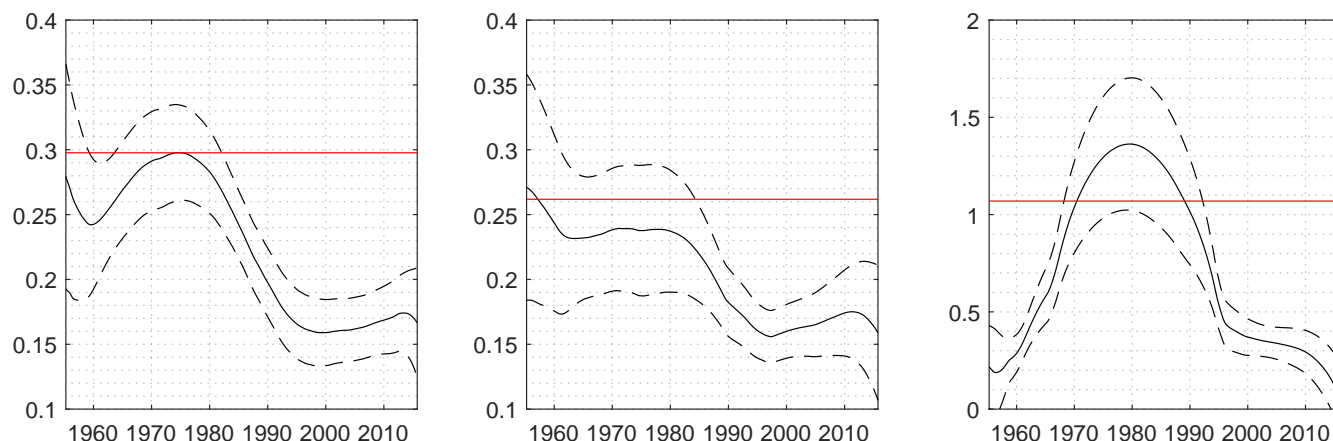


Figure 3: The standard deviation of estimation residuals $\hat{\eta}_t$ in the inflation equation (left), the unemployment equation (middle) and the interest rate equation (right) as well as the associated 95% confidence intervals. The red line denotes the estimated standard deviation using the VAR(3) model with constant parameters.

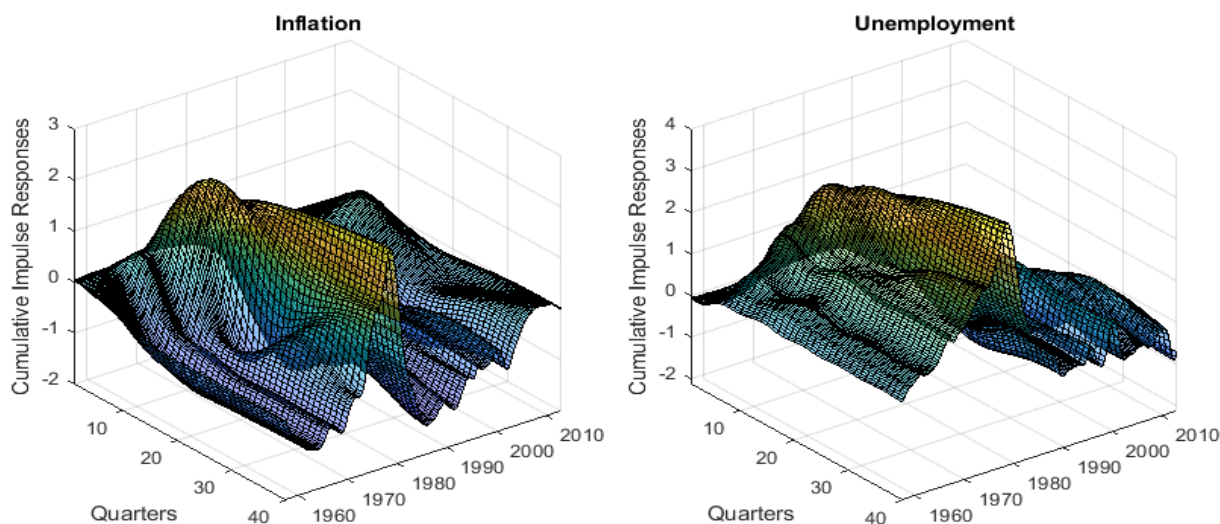


Figure 4: Time-varying cumulative impulse responses to the monetary policy shocks. Left front axis left: Quarters (Horizon). Right front axis left: Time.

Additionally, Figure 5 plots the impulse responses of inflation to a monetary policy shock in 1978:Q1 and 1994:Q1. The left panel in Figure 5 exhibits the price puzzle, while the right panel shows a response that corresponds to the mainstream prior: the price level declines soon after a tightening. This result shows that the price puzzle is limited to periods of bad monetary policy, and is consistent with a stream of literature (e.g., Elbourne and de Haan, 2006; Rusnák et al., 2013), which suggests that the price puzzle emerges when using data for different monetary policy regimes and VARs estimated within a period of single monetary policy rarely encounter the price puzzle. Hence, the existence of time-varying policy transmission mechanism favours the TV-VAR(3) model.

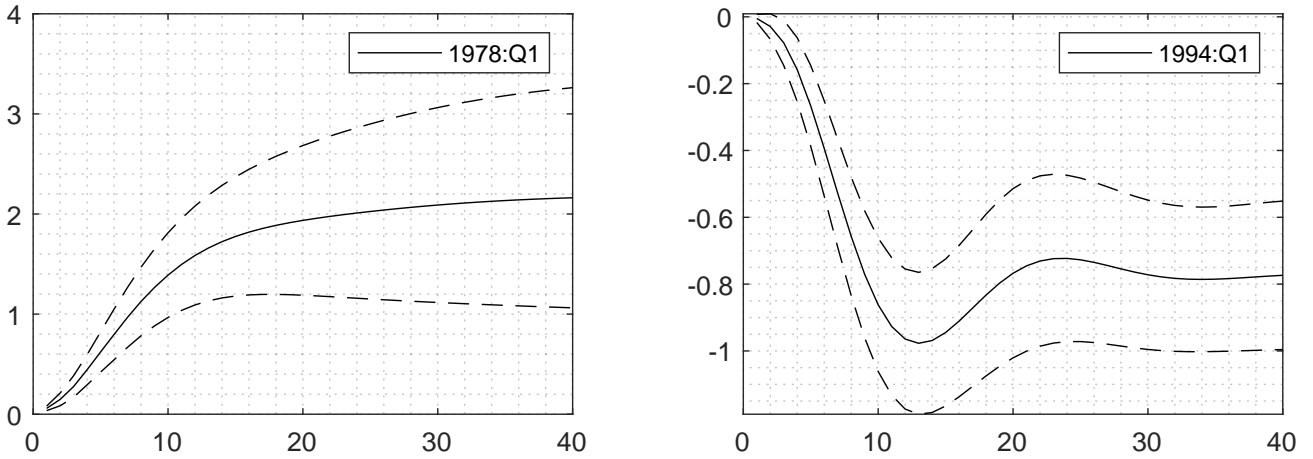


Figure 5: Impulse responses of inflation to the monetary policy shocks in 1978:Q1 and 1994:Q1 with one-standard error confidence intervals.

6 Conclusions and Discussion

In this paper, we have proposed a new class of time-varying VAR models where the VAR coefficients and covariance matrix of the error innovations are allowed to evolve over time. Accordingly, we have established a set of asymptotic results, including the impulse responses analyses subject to both short-run timing and long-run restrictions, an information criterion to select the optimal lag, and a Wald-type test to determine the constant coefficients. Simulation studies are conducted to evaluate the theoretical findings. Finally, we demonstrate the empirical relevance and usefulness of the proposed methods through an application to the transmission mechanism of U.S. monetary policy. We have revealed that there exists a substantial time-variation in the policy transmission mechanism, and the “price puzzle” is limited to periods of bad monetary policy.

There are several directions for possible extensions. The first one is about how to consistently estimate the d -dimensional components of the VAR(p) process for the case where the dimensionality, d , and the number of lags, p , may diverge along with the sample size, T . The second one is to allow for some time-varying structure in cointegrated dynamic models. We wish to leave such issues for future study.

Appendix A

For the sake of presentation, we first provide some notations and mathematical symbols in Appendix A.1. The proofs of main results are provided in Appendix A.2. In what follows, M and $O(1)$ always stand for constants, and may be different at each appearance.

A.1 Notations and Mathematical Symbols

For ease of notation, we define three matrices $\boldsymbol{\Sigma}(\tau)$, $\mathbf{V}(\tau)$ and $\boldsymbol{\Phi}(\tau)$ with their estimators respectively. For $\forall \tau \in (0, 1)$, let

$$\boldsymbol{\Sigma}(\tau) = \begin{bmatrix} 1 & \boldsymbol{\mu}^\top(\tau) & \cdots & \boldsymbol{\mu}^\top(\tau) \\ \boldsymbol{\mu}(\tau) & \boldsymbol{\Sigma}_0(\tau) & \cdots & \boldsymbol{\Sigma}_{p-1}^\top(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu}(\tau) & \boldsymbol{\Sigma}_{p-1}(\tau) & \cdots & \boldsymbol{\Sigma}_0(\tau) \end{bmatrix}, \quad (\text{A.1})$$

in which $\boldsymbol{\mu}(\tau)$ and $\mathbf{B}_j(\tau)$ are defined in Proposition 2.1 and $\boldsymbol{\Sigma}_m(\tau) = \boldsymbol{\mu}(\tau)\boldsymbol{\mu}(\tau)^\top + \sum_{j=0}^{\infty} \mathbf{B}_j(\tau)\mathbf{B}_{j+m}^\top(\tau)$ for $m = 0, \dots, p-1$. We define the estimator of $\boldsymbol{\Sigma}(\tau)$ as

$$\widehat{\boldsymbol{\Sigma}}(\tau) = \left(\frac{1}{T} \sum_{t=1}^T K_h(\tau_t - \tau) \right)^{-1} \frac{1}{T} \sum_{t=1}^T \mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top K_h(\tau_t - \tau), \quad (\text{A.2})$$

where \mathbf{z}_t is defined in (2.4).

Next, we let

$$\mathbf{V}(\tau) = \begin{bmatrix} \mathbf{V}_{1,1}(\tau) & \mathbf{V}_{2,1}^\top(\tau) \\ \mathbf{V}_{2,1}(\tau) & \mathbf{V}_{2,2}(\tau) \end{bmatrix}, \quad (\text{A.3})$$

where $\mathbf{V}_{1,1}(\tau) = \tilde{v}_0 \boldsymbol{\Sigma}^{-1}(\tau) \otimes \boldsymbol{\Omega}(\tau)$,

$$\begin{aligned} \mathbf{V}_{2,1}(\tau) &= \lim_{T \rightarrow \infty} \frac{h}{T} \sum_{t=1}^T E \left(\text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top) \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \right) K_h(\tau_t - \tau)^2 \cdot (\boldsymbol{\Sigma}^{-1}(\tau) \otimes \mathbf{I}_d), \\ \mathbf{V}_{2,2}(\tau) &= \lim_{T \rightarrow \infty} \frac{h}{T} \sum_{t=1}^T E \left(\text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top) \text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top)^\top \right) K_h(\tau_t - \tau)^2 \\ &\quad - \tilde{v}_0 \text{vech}(\boldsymbol{\Omega}(\tau)) \text{vech}(\boldsymbol{\Omega}(\tau))^\top. \end{aligned} \quad (\text{A.4})$$

The estimator of $\mathbf{V}(\tau)$ is then defined as follows:

$$\widehat{\mathbf{V}}(\tau) = \begin{bmatrix} \widehat{\mathbf{V}}_{1,1}(\tau) & \widehat{\mathbf{V}}_{2,1}^\top(\tau) \\ \widehat{\mathbf{V}}_{2,1}(\tau) & \widehat{\mathbf{V}}_{2,2}(\tau) \end{bmatrix}, \quad (\text{A.5})$$

where $\widehat{\mathbf{V}}_{1,1}(\tau)$, $\widehat{\mathbf{V}}_{2,1}(\tau)$ and $\widehat{\mathbf{V}}_{2,2}(\tau)$ have the forms identical to their counterparts of (A.4), but we replace $\boldsymbol{\Sigma}(\tau)$, $\boldsymbol{\eta}_t$ and $\boldsymbol{\Omega}(\tau)$ with their estimators presented in (A.2) and (2.5).

Finally, recall the following definition:

$$\Phi(\tau) = \begin{bmatrix} \mathbf{A}_1(\tau) & \cdots & \mathbf{A}_{p-1}(\tau) & \mathbf{A}_p(\tau) \\ \mathbf{I}_d & \cdots & \mathbf{0}_d & \mathbf{0}_d \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_d & \cdots & \mathbf{I}_d & \mathbf{0}_d \end{bmatrix}. \quad (\text{A.6})$$

Replacing $\mathbf{A}_j(\tau)$'s of (A.6) with their estimators obtained from (2.5) yields an estimator, $\widehat{\Phi}(\tau)$, for $\Phi(\tau)$.

A.2 Proofs of the Main Results

Proof of Proposition 2.1.

Consider the VMA representation of \mathbf{x}_t : $\mathbf{x}_t = \boldsymbol{\mu}_t + \mathbf{B}_{0,t}\boldsymbol{\epsilon}_t + \mathbf{B}_{1,t}\boldsymbol{\epsilon}_{t-1} + \mathbf{B}_{2,t}\boldsymbol{\epsilon}_{t-2} + \cdots$, where $\mathbf{B}_{0,t} = \boldsymbol{\omega}(\tau_t)$, $\mathbf{B}_{j,t} = \boldsymbol{\Psi}_{j,t}\boldsymbol{\omega}(\tau_{t-j})$, $\boldsymbol{\Psi}_{j,t} = \mathbf{J} \prod_{m=0}^{j-1} \Phi(\tau_{t-m}) \mathbf{J}^\top$ for $j \geq 1$, $\boldsymbol{\mu}_t = \mathbf{a}(\tau_t) + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_{j,t} \mathbf{a}(\tau_{t-j})$ and $\tau_{t-j} = \frac{t-j}{T} I(t \geq j)$.

First, we investigate the validity of the VMA representations of \mathbf{x}_t and $\tilde{\mathbf{x}}_t$. Let ρ_A denote the largest eigenvalue of $\Phi(\tau)$ uniformly over $\tau \in [0, 1]$. Then, $\rho_A < 1$ by Assumption 4.1. Similar to the proof of Proposition 2.4 in Dahlhaus and Polonik (2009), we have $\max_{t \geq 1} \left\| \prod_{m=0}^{j-1} \Phi(\tau_{t-m}) \right\| \leq M \rho_A^j$. It follows that $\|\mathbf{E}(\mathbf{x}_t)\| \leq \sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_{j,t}\| \cdot \|\mathbf{a}(\tau_{t-j})\| \leq M \sum_{j=0}^{\infty} \rho_A^j < \infty$ and

$$\|\text{Var}(\mathbf{x}_t)\| = \left\| \sum_{j=0}^{\infty} \mathbf{B}_{j,t} \mathbf{B}_{j,t}^\top \right\| \leq \sum_{j=0}^{\infty} \|\mathbf{B}_{j,t}\|^2 \leq M \sum_{j=0}^{\infty} \rho_A^{2j} < \infty.$$

Similarly, we have $\|\mathbf{E}(\tilde{\mathbf{x}}_t)\| < \infty$ and $\|\text{Var}(\tilde{\mathbf{x}}_t)\| < \infty$.

Then, we need to verify that $\max_{t \geq 1} \{E \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|^\delta\}^{1/\delta} = O(T^{-1})$. For any conformable matrices $\{\mathbf{A}_i\}$ and $\{\mathbf{B}_i\}$, since $\prod_{i=1}^r \mathbf{A}_i - \prod_{i=1}^r \mathbf{B}_i = \sum_{j=1}^r \left(\prod_{k=1}^{j-1} \mathbf{A}_k \right) (\mathbf{A}_j - \mathbf{B}_j) \left(\prod_{k=j+1}^r \mathbf{B}_k \right)$, we then obtain

$$\begin{aligned} \|\mathbf{B}_{j,t} - \mathbf{B}_j(\tau_t)\| &= \left\| \mathbf{J} \prod_{m=0}^{j-1} \Phi(\tau_{t-m}) \mathbf{J}^\top \boldsymbol{\omega}(\tau_{t-j}) - \mathbf{J} \Phi^j(\tau_t) \mathbf{J}^\top \boldsymbol{\omega}(\tau_t) \right\| \\ &= \left\| \left(\mathbf{J} \prod_{m=0}^{j-1} \Phi(\tau_{t-m}) \mathbf{J}^\top - \mathbf{J} \Phi^j(\tau_t) \mathbf{J}^\top \right) \boldsymbol{\omega}(\tau_t) + \mathbf{J} \prod_{m=0}^{j-1} \Phi(\tau_{t-m}) \mathbf{J}^\top (\boldsymbol{\omega}(\tau_{t-j}) - \boldsymbol{\omega}(\tau_t)) \right\| \\ &\leq M \sum_{i=1}^{j-1} \left\| \Phi^i(\tau_t) (\Phi(\tau_{t-i}) - \Phi(\tau_t)) \prod_{m=i+1}^{j-1} \Phi(\tau_{t-m}) \right\| + M \rho_A^j \frac{j}{T} \leq M \sum_{i=1}^{j-1} \frac{i}{T} \rho_A^{j-1} + M \rho_A^j \frac{j}{T} = O(T^{-1}), \end{aligned}$$

which implies for the same $\delta > 4$ as in Assumption 2,

$$\begin{aligned} \{E \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|^\delta\}^{1/\delta} &\leq \sum_{j=1}^{\infty} \|\boldsymbol{\Psi}_{j,t} \mathbf{a}(\tau_{t-j}) - \boldsymbol{\Psi}_j(\tau_t) \mathbf{a}(\tau_t)\| + \sum_{j=1}^{\infty} \|\mathbf{B}_{j,t} - \mathbf{B}_j(\tau_t)\| \cdot \{E \|\boldsymbol{\epsilon}_t\|^\delta\}^{1/\delta} \\ &\leq M \sum_{j=1}^{\infty} \left(\sum_{i=1}^{j-1} \frac{i}{T} \rho_A^{j-1} + \rho_A^j \frac{j}{T} \right) = O(T^{-1}). \end{aligned}$$

The proof is now completed. □

Proof of Theorem 2.1.

(1). For notation simplicity, let $\mathbf{S}_{T,k}(\tau) = \frac{1}{T} \sum_{t=1}^T \mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top \left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau)$,

$$\mathbf{S}_T(\tau) = \begin{pmatrix} \mathbf{S}_{T,0}(\tau) & \mathbf{S}_{T,1}(\tau) \\ \mathbf{S}_{T,1}(\tau) & \mathbf{S}_{T,2}(\tau) \end{pmatrix},$$

and $\mathbf{M}(\tau_t) = \mathbf{A}(\tau_t) - \mathbf{A}(\tau) - \mathbf{A}^{(1)}(\tau)(\tau_t - \tau) - \frac{1}{2} \mathbf{A}^{(2)}(\tau)(\tau_t - \tau)^2$.

We now begin our investigation. Since

$$\begin{aligned} \mathbf{x}_t &= \left(\mathbf{A}(\tau) + \mathbf{A}^{(1)}(\tau)(\tau_t - \tau) + \frac{1}{2} \mathbf{A}^{(2)}(\tau)(\tau_t - \tau)^2 + \mathbf{M}(\tau_t) \right) \mathbf{z}_{t-1} + \boldsymbol{\eta}_t \\ &= \mathbf{Z}_{t-1}^{*,\top} \begin{bmatrix} \text{vec}(\mathbf{A}(\tau)) \\ h \text{vec}(\mathbf{A}^{(1)}(\tau)) \end{bmatrix} + \frac{1}{2} h^2 \left(\frac{\tau_t - \tau}{h}\right)^2 (\mathbf{z}_{t-1}^\top \otimes \mathbf{I}_d) \text{vec}(\mathbf{A}^{(2)}(\tau)) \\ &\quad + (\mathbf{z}_{t-1}^\top \otimes \mathbf{I}_d) \text{vec}(\mathbf{M}(\tau_t)) + \boldsymbol{\eta}_t, \end{aligned}$$

we write

$$\begin{aligned} &\text{vec}(\widehat{\mathbf{A}}(\tau) - \mathbf{A}(\tau)) \\ &= [\mathbf{I}_{d^2 p+d}, \mathbf{0}_{d^2 p+d}] \cdot \left(\frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{t-1}^* \mathbf{Z}_{t-1}^{*,\top} K_h(\tau_t - \tau) \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{t-1}^* \mathbf{x}_t K_h(\tau_t - \tau) \right) - \text{vec}(\mathbf{A}(\tau)) \\ &= [\mathbf{I}_{d^2 p+d}, \mathbf{0}_{d^2 p+d}] \left(\mathbf{S}_T^{-1}(\tau) \begin{pmatrix} \mathbf{S}_{T,2}(\tau) \\ \mathbf{S}_{T,3}(\tau) \end{pmatrix} \otimes \mathbf{I}_d \right) \left\{ \frac{1}{2} h^2 \text{vec}(\mathbf{A}^{(2)}(\tau)) \right\} \\ &\quad + [\mathbf{I}_{d^2 p+d}, \mathbf{0}_{d^2 p+d}] (\mathbf{S}_T^{-1}(\tau) \otimes \mathbf{I}_d) \left(\frac{1}{T} \sum_{t=1}^T (\mathbf{z}_{t-1}^* \mathbf{z}_{t-1}^\top \otimes \mathbf{I}_d) \text{vec}(\mathbf{M}(\tau_t)) K_h(\tau_t - \tau) \right) \\ &\quad + [\mathbf{I}_{d^2 p+d}, \mathbf{0}_{d^2 p+d}] (\mathbf{S}_T^{-1}(\tau) \otimes \mathbf{I}_d) \left(\frac{1}{T} \sum_{t=1}^T (\mathbf{z}_{t-1}^* \otimes \mathbf{I}_d) \boldsymbol{\eta}_t K_h(\tau_t - \tau) \right) \\ &:= I_{T,1} + I_{T,2} + I_{T,3}. \end{aligned}$$

By standard arguments for the local linear kernel estimator and the uniform convergence results in Lemmas B.5.2-3, we have $\|I_{T,1} + I_{T,2}\| = O(h^2) + O_P(h^2 \sqrt{\log T / (Th)})$ uniformly over $\tau \in [0, 1]$. By Lemma B.6.1, we have $I_{T,3} = O_P\left(\left(\frac{\log T}{Th}\right)^{\frac{1}{2}}\right)$ uniformly over $\tau \in [0, 1]$. Therefore, the first result follows.

(2). We begin our investigation on the asymptotic normality by writing that for $\forall \tau \in (0, 1)$,

$$\begin{aligned} \widehat{\boldsymbol{\Omega}}(\tau) &= \frac{1}{Th} \sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) + O_P\left(\frac{1}{Th}\right) \\ &= \frac{1}{Th} \sum_{t=1}^T (\boldsymbol{\eta}_t + \widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t) (\boldsymbol{\eta}_t + \widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) + O_P\left(\frac{1}{Th}\right) \\ &= \frac{1}{Th} \sum_{t=1}^T \boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) + \frac{1}{Th} \sum_{t=1}^T (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t) (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) \\ &\quad + \frac{1}{Th} \sum_{t=1}^T \boldsymbol{\eta}_t (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) + \frac{1}{Th} \sum_{t=1}^T (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t) \boldsymbol{\eta}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) + O_P\left(\frac{1}{Th}\right) \end{aligned}$$

$$:= \mathbf{I}_{T,4} + \mathbf{I}_{T,5} + \mathbf{I}_{T,6} + \mathbf{I}_{T,7} + O_P\left(\frac{1}{Th}\right).$$

Let $\rho_T = h^2 + \sqrt{\frac{\log T}{Th}}$. By what we have just proved for Theorem 2.1(i), for $\forall \tau \in [0, 1]$ we have

$$\begin{aligned} & \left\| \frac{1}{Th} \sum_{t=1}^T (\hat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) \right\| \\ & \leq \sup_{\tau_t \in [0,1]} \|\hat{\mathbf{A}}(\tau_t) - \mathbf{A}(\tau_t)\|^2 \cdot \frac{1}{Th} \sum_{t=1}^T \|\mathbf{z}_{t-1}\|^2 K\left(\frac{\tau_t - \tau}{h}\right) = O_P(\rho_T^2). \end{aligned}$$

By Lemma B.6.2, $\mathbf{I}_{T,6}$ and $\mathbf{I}_{T,7}$ are both $o_P((Th)^{-1/2})$. Hence,

$$\sqrt{Th} \left(\frac{1}{Th} \sum_{t=1}^T \hat{\boldsymbol{\eta}}_t \hat{\boldsymbol{\eta}}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) - \frac{1}{Th} \sum_{t=1}^T \boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) - o_P(h^4) \right) = o_P(1).$$

Combined with the convergence results of the sample covariance matrix stated in Lemma B.5, the above development yields that

$$\begin{aligned} & \sqrt{Th} \begin{bmatrix} \text{vec} \left(\hat{\mathbf{A}}(\tau) - \mathbf{A}(\tau) - \frac{1}{2} h^2 \tilde{c}_2 \mathbf{A}^{(2)}(\tau) \right) + o_P(h^2) \\ \text{vech} \left(\hat{\boldsymbol{\Omega}}(\tau) - \boldsymbol{\Omega}(\tau) - \frac{1}{2} h^2 \tilde{c}_2 \boldsymbol{\Omega}^{(2)}(\tau) \right) + o_P(h^2) \end{bmatrix} \\ & = \begin{bmatrix} (\boldsymbol{\Sigma}^{-1}(\tau) \otimes \mathbf{I}_d) \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{z}_{t-1} \boldsymbol{\eta}_t K\left(\frac{\tau_t - \tau}{h}\right) \right) \\ \frac{1}{\sqrt{Th}} \sum_{t=1}^T \text{vech} \left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top - \boldsymbol{\Omega}(\tau_t) \right) K\left(\frac{\tau_t - \tau}{h}\right) \end{bmatrix} + o_P(1) := \mathbf{I}_{T,8} + o_P(1). \end{aligned}$$

Below, we focus on $\mathbf{I}_{T,8}$. First, we show $\text{Var}(\mathbf{I}_{T,8}) \rightarrow \mathbf{V}(\tau)$. Let $\text{Var}(\mathbf{I}_{T,8}) = \begin{bmatrix} \tilde{\mathbf{V}}_{1,1}(\tau) & \tilde{\mathbf{V}}_{2,1}^\top(\tau) \\ \tilde{\mathbf{V}}_{2,1}(\tau) & \tilde{\mathbf{V}}_{2,2}(\tau) \end{bmatrix}$, where

the definition of each block should be obvious. Moreover, simple algebra shows that $\tilde{\mathbf{V}}_{i,j}(\tau) \rightarrow \mathbf{V}_{i,j}(\tau)$ for $i, j \in \{1, 2\}$.

By construction and Assumption 2, $\mathbf{I}_{T,8}$ is a summation of m.d.s., we thus use Lemma B.1 and Cramér-Wold device to prove its asymptotic normality. It suffices to show that $\mathbf{d}^\top \mathbf{I}_{T,8} \rightarrow_D N(\mathbf{0}, \mathbf{d}^\top \mathbf{V}(\tau) \mathbf{d})$ for any conformable unit vector \mathbf{d} . Let

$$\mathbf{Z}_{T,t}(\tau) = \frac{1}{\sqrt{Th}} \mathbf{d}^\top \begin{bmatrix} (\boldsymbol{\Sigma}^{-1}(\tau) \otimes \mathbf{I}_d) \left(\mathbf{z}_{t-1} \boldsymbol{\eta}_t K\left(\frac{\tau_t - \tau}{h}\right) \right) \\ \text{vech} \left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top - \boldsymbol{\Omega}(\tau_t) \right) K\left(\frac{\tau_t - \tau}{h}\right) \end{bmatrix}.$$

By the law of large numbers for martingale differences, we have

$$\sum_{t=1}^T \mathbf{Z}_{T,t}^2(\tau) - \sum_{t=1}^T E(\mathbf{Z}_{T,t}^2(\tau) | \mathcal{F}_{t-1}) \rightarrow_P 0.$$

Since conditional on \mathcal{F}_{t-1} the third and fourth moments of $\boldsymbol{\epsilon}_t$ are identical to the corresponding unconditional moments a.s., by Lemma B.5.1 we can prove that $\sum_{t=1}^T E(\mathbf{Z}_{T,t}^2(\tau) | \mathcal{F}_{t-1}) \rightarrow_P \mathbf{d}^\top \mathbf{V}(\tau) \mathbf{d}$.

Furthermore, for any $\nu > 0$ and $\tau \in (0, 1)$, by both Holder's and Markov's inequalities, we have

$$\begin{aligned} \sum_{t=1}^T E \left((\mathbf{Z}_{T,t}(\tau))^2 I(|\mathbf{Z}_{T,t}(\tau)| > \nu) \right) &\leq \sum_{t=1}^T \left[E|\mathbf{Z}_{T,t}(\tau)|^{\delta/2} \right]^{4/\delta} \left[\frac{E|\mathbf{Z}_{T,t}(\tau)|^{\delta/2}}{\nu^{\delta/2}} \right]^{(\delta-4)/\delta} \\ &= O((Th)^{(\delta-4)/4}) = o(1). \end{aligned}$$

Thus, the CLT follows.

Finally, we consider $\widehat{\mathbf{V}}(\cdot)$. By Lemma B.5 and the above proof, we have $\widehat{\mathbf{V}}_{1,1}(\tau) \rightarrow_P \mathbf{V}_{1,1}(\tau)$. By the uniform convergence results of $\widehat{\mathbf{A}}(\tau)$, we can replace $\widehat{\boldsymbol{\eta}}_t$ with $\boldsymbol{\eta}_t$ in the following derivations. Therefore, we have

$$\widehat{\mathbf{V}}_{2,1}(\tau) = \frac{1}{Th} \sum_{t=1}^T \text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top) \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top K^2 \left(\frac{\tau_t - \tau}{h} \right) (\boldsymbol{\Sigma}^{-1}(\tau) \otimes \mathbf{I}_d) + o_P(1) \rightarrow_P \mathbf{V}_{2,1}(\tau),$$

and

$$\begin{aligned} \widehat{\mathbf{V}}_{2,2}(\tau) &= \frac{1}{Th} \sum_{t=1}^T \text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top) \text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top)^\top K^2 \left(\frac{\tau_t - \tau}{h} \right) - \tilde{v}_0 \text{vech}(\boldsymbol{\Omega}(\tau)) \text{vech}(\boldsymbol{\Omega}(\tau))^\top + o_P(1) \\ &\rightarrow_P \mathbf{V}_{2,2}(\tau). \end{aligned}$$

The proof is now completed. □

Proof of Theorem 2.2.

Let $\mathbf{A}(\boldsymbol{\theta})$ be a real, differentiable, $m \times n$ matrix function of real $p \times 1$ vector $\boldsymbol{\theta}$. Define $\nabla_{\boldsymbol{\theta}} \mathbf{A} = \frac{\partial \text{vec}(\mathbf{A})}{\partial \boldsymbol{\theta}^\top}$, and thus $\text{vec}(d\mathbf{A}) = \nabla_{\boldsymbol{\theta}} \mathbf{A} d\boldsymbol{\theta}$.

Let $\boldsymbol{\alpha}(\tau) = \text{vec}(\mathbf{A}_1(\tau), \dots, \mathbf{A}_p(\tau))$, $\boldsymbol{\sigma}(\tau) = \text{vech}(\boldsymbol{\Omega}(\tau))$ and $\boldsymbol{\phi}(\tau) = [\boldsymbol{\alpha}^\top(\tau), \boldsymbol{\sigma}^\top(\tau)]^\top$. Given the joint distribution of $\text{vec}(\widehat{\mathbf{A}}(\tau))$ and $\text{vech}(\widehat{\boldsymbol{\Omega}}(\tau))$ in Theorem 2.1, Theorem 2.2 can be obtained by the Delta method. By the first-order approximation of $\text{vec}(\widehat{\mathbf{B}}_j(\tau))$ around $\text{vec}(\mathbf{B}_j(\tau))$, we have

$$\sqrt{Th} \text{vec}(\widehat{\mathbf{B}}_j(\tau) - \mathbf{B}_j(\tau)) \simeq \nabla_{\boldsymbol{\phi}(\tau)} \mathbf{B}_j(\tau) \sqrt{Th} (\widehat{\boldsymbol{\phi}}(\tau) - \boldsymbol{\phi}(\tau))$$

and thus

$$\sqrt{Th} \left(\text{vec}(\widehat{\mathbf{B}}_j(\tau) - \mathbf{B}_j(\tau)) - \frac{1}{2} h^2 \tilde{c}_2 \mathbf{B}_j^{(2)}(\tau) + o_P(h^2) \right) \rightarrow_D N(0, \boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)),$$

where $\mathbf{B}_j^{(2)}(\tau)$ and $\boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)$ have been defined in the body of the theorem.

To complete the proof, we first derive an analytic form for the derivative $\nabla_{\boldsymbol{\phi}(\tau)} \mathbf{B}_j(\tau)$ under each of the identification restrictions.

We have two sets of restrictions: (a) $d(d+1)/2$ restrictions implied by $\boldsymbol{\Omega}(\tau) = \boldsymbol{\omega}(\tau) \boldsymbol{\omega}^\top(\tau)$ and (b) additional $d(d-1)/2$ structural restrictions based on short-run or long-run restrictions.

Consider type (a) restrictions. We begin by considering $d\boldsymbol{\Omega}(\tau) = d\boldsymbol{\omega}(\tau) \cdot \boldsymbol{\omega}^\top(\tau) + \boldsymbol{\omega}(\tau) \cdot d\boldsymbol{\omega}^\top(\tau)$. Let \mathbf{B} and \mathbf{C} be $n \times q$ and $q \times r$ matrices, respectively. By $\text{vec}(\mathbf{ABC}) = \mathbf{C}^\top \otimes \mathbf{A} \text{vec}(\mathbf{B})$, $\text{vec}(\mathbf{A}^\top) = \mathbf{K}_{m,n} \text{vec}(\mathbf{A})$ and $\mathbf{K}_{m,q}(\mathbf{A} \otimes \mathbf{C}) = (\mathbf{C} \otimes \mathbf{A}) \mathbf{K}_{n,r}$, we have $\mathbf{N}_1(\tau) \text{vec}(d\boldsymbol{\omega}(\tau)) = \text{vec}(d\boldsymbol{\Omega}(\tau))$, where $\mathbf{N}_1(\tau) = (\mathbf{I}_{d^2} + \mathbf{K}_{d,d})(\boldsymbol{\omega}(\tau) \otimes \mathbf{I}_d)$. Let \mathbf{D}_1 be the duplication matrix such that $\text{vec}[\boldsymbol{\Omega}(\tau)] = \mathbf{D}_1 \text{vech}[\boldsymbol{\Omega}(\tau)]$, which follows

that $\mathbf{N}_1(\tau)\text{vec}(d\boldsymbol{\omega}(\tau)) = \mathbf{D}_1 d\boldsymbol{\sigma}(\tau)$ and

$$\mathbf{N}_1(\tau)\nabla_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau) = \mathbf{D}_1. \quad (\text{A.1})$$

We then illustrate how to combine equation (A.1) with gradient equations from type (b) restrictions in order to compute $\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$.

In the case of short-run timing restrictions, because types (a) and (b) restrictions do not involve $\boldsymbol{\alpha}$, $\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$ has the form $[\mathbf{0}, \nabla_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]$. Let \mathbf{L}_d be the elimination matrix defined by $\text{vech}[\boldsymbol{\omega}(\tau)] = \mathbf{L}_d \text{vec}[\boldsymbol{\omega}(\tau)]$. Because $\boldsymbol{\omega}(\tau)$ is lower triangular subject to short-run restrictions, \mathbf{L}_d^\top is a duplication matrix such that $\text{vec}[\boldsymbol{\omega}(\tau)] = \mathbf{L}_d^\top \text{vech}[\boldsymbol{\omega}(\tau)]$. Write

$$\begin{aligned} \mathbf{N}_1(\tau)\text{vec}(d\boldsymbol{\omega}(\tau)) &= \mathbf{D}_1 d\boldsymbol{\sigma}(\tau), \\ \mathbf{L}_d \mathbf{N}_1(\tau) \mathbf{L}_d^\top \text{vech}(d\boldsymbol{\omega}(\tau)) &= \mathbf{L}_d \mathbf{D}_1 d\boldsymbol{\sigma}(\tau) = d\boldsymbol{\sigma}(\tau), \\ \text{vech}(d\boldsymbol{\omega}(\tau)) &= \left(\mathbf{L}_d \mathbf{N}_1(\tau) \mathbf{L}_d^\top \right)^{-1} d\boldsymbol{\sigma}(\tau), \\ \text{vec}(d\boldsymbol{\omega}(\tau)) &= \mathbf{L}_d^\top \left(\mathbf{L}_d \mathbf{N}_1(\tau) \mathbf{L}_d^\top \right)^{-1} d\boldsymbol{\sigma}(\tau). \end{aligned}$$

Hence, $\nabla_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau) = \mathbf{L}_d^\top \left(\mathbf{L}_d \mathbf{N}_1(\tau) \mathbf{L}_d^\top \right)^{-1}$. Recall that $\nabla_{\boldsymbol{\phi}(\tau)}\mathbf{B}_j(\tau) = [\nabla_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_j(\tau), \nabla_{\boldsymbol{\sigma}(\tau)}\mathbf{B}_j(\tau)]$. For $\nabla_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_j(\tau)$,

$$\begin{aligned} \nabla_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_j(\tau) &= \frac{\partial \text{vec}[\boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}(\tau)]}{\partial \boldsymbol{\alpha}^\top(\tau)} = (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_d) \frac{\partial \text{vec}[\boldsymbol{\Psi}_j(\tau)]}{\partial \boldsymbol{\alpha}^\top(\tau)} \\ &= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_d) \frac{\partial \text{vec}[\mathbf{J}\boldsymbol{\Phi}^j(\tau)\mathbf{J}^\top]}{\partial \boldsymbol{\alpha}^\top(\tau)} \\ &= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_d) \left(\sum_{m=0}^{j-1} \mathbf{J}(\boldsymbol{\Phi}^\top(\tau))^{j-1-m} \otimes \boldsymbol{\Psi}_m(\tau) \right). \end{aligned}$$

For $\nabla_{\boldsymbol{\sigma}(\tau)}\mathbf{B}_j(\tau)$,

$$\begin{aligned} \nabla_{\boldsymbol{\sigma}(\tau)}\mathbf{B}_j(\tau) &= \frac{\partial \text{vec}[\boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}(\tau)]}{\partial \boldsymbol{\sigma}^\top(\tau)} = (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \frac{\partial \text{vec}[\boldsymbol{\omega}(\tau)]}{\partial \boldsymbol{\sigma}^\top(\tau)} \\ &= (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \mathbf{L}_d^\top \left(\mathbf{L}_d \mathbf{N}_1(\tau) \mathbf{L}_d^\top \right)^{-1}. \end{aligned}$$

In the case of long-run restrictions, type (b) restrictions involve $\boldsymbol{\alpha}(\tau)$, so that $\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$ has the form $[\nabla_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\omega}(\tau), \nabla_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]$. First, equation (A.1) must be extended in the form $\mathbf{N}_1 \nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) = [\mathbf{0}, \mathbf{D}_1]$. Second, long-run restrictions can be expressed as $\mathbf{Q} \text{vec}[\mathbf{A}_\tau^{-1}(1)\boldsymbol{\omega}] = 0$, where \mathbf{Q} is a $d(d-1)/2 \times d^2$ matrix of 0 and 1, and $\mathbf{A}_\tau(1) = \mathbf{I}_d - \sum_{i=1}^p \mathbf{A}_i(\tau)$. By $d\mathbf{A}^{-1} = -\mathbf{A}^{-1} \cdot d\mathbf{A} \cdot \mathbf{A}^{-1}$, we have

$$\begin{aligned} \mathbf{Q} \text{vec}[\mathbf{A}_\tau^{-1}(1)\boldsymbol{\omega}] &= 0 \\ \mathbf{Q} \text{vec}[d(\mathbf{A}_\tau^{-1}(1))\boldsymbol{\omega} + \mathbf{A}_\tau^{-1}(1)d\boldsymbol{\omega}] &= 0 \\ \mathbf{Q} \text{vec}[-\mathbf{A}_\tau^{-1}(1)d(\mathbf{A}_\tau(1))\mathbf{A}_\tau^{-1}(1)\boldsymbol{\omega} + \mathbf{A}_\tau^{-1}(1)d\boldsymbol{\omega}] &= 0 \\ \mathbf{Q} [\mathbf{I}_d \otimes \mathbf{A}_\tau^{-1}(1)] \text{vec}[d\boldsymbol{\omega}] &= \mathbf{Q}[\mathbf{B}^\top(\tau) \otimes \mathbf{A}_\tau^{-1}(1)] \text{vec}[d\mathbf{A}_\tau(1)] \\ \mathbf{N}_2(\tau)\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) &= [\mathbf{D}_2(\tau), \mathbf{0}], \end{aligned}$$

where $\mathbf{N}_2(\tau) = \mathbf{Q} [\mathbf{I}_d \otimes \mathbf{A}_\tau^{-1}(1)]$ and $\mathbf{D}_2(\tau) = \mathbf{Q}[\mathbf{B}^\top(\tau) \otimes \mathbf{A}_\tau^{-1}(1)] \nabla_{\boldsymbol{\alpha}(\tau)} \mathbf{A}_\tau(1)$ with $\nabla_{\boldsymbol{\alpha}(\tau)} \mathbf{A}_\tau(1) = -[\mathbf{I}_{d^2}, \dots, \mathbf{I}_{d^2}]$ ($d^2 \times d^2 p$). Hence,

$$\begin{aligned} \nabla_{\phi(\tau)} \boldsymbol{\omega}(\tau) &= [\nabla_{\boldsymbol{\alpha}(\tau)} \boldsymbol{\omega}(\tau), \nabla_{\boldsymbol{\sigma}(\tau)} \boldsymbol{\omega}(\tau)] \\ &= \left[\left(\mathbf{N}_1^\top(\tau), \mathbf{N}_2^\top(\tau) \right) \begin{pmatrix} \mathbf{N}_1(\tau) \\ \mathbf{N}_2(\tau) \end{pmatrix} \right]^{-1} \left[\mathbf{N}_1^\top(\tau), \mathbf{N}_2^\top(\tau) \right] \begin{bmatrix} \mathbf{0} & \mathbf{D}_1 \\ \mathbf{D}_2(\tau) & \mathbf{0} \end{bmatrix} \\ &= \left(\mathbf{N}_1^\top(\tau) \mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau) \mathbf{N}_2(\tau) \right)^{-1} \left[\mathbf{N}_2^\top(\tau) \mathbf{D}_2(\tau), \mathbf{N}_1^\top(\tau) \mathbf{D}_1 \right]. \end{aligned}$$

For $\nabla_{\boldsymbol{\alpha}(\tau)} \mathbf{B}_j(\tau)$,

$$\begin{aligned} \nabla_{\boldsymbol{\alpha}(\tau)} \mathbf{B}_j(\tau) &= \frac{\partial \text{vec} [\boldsymbol{\Psi}_j(\tau) \boldsymbol{\omega}(\tau)]}{\partial \boldsymbol{\alpha}^\top(\tau)} \\ &= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_d) \frac{\partial \text{vec} [\boldsymbol{\Psi}_j(\tau)]}{\partial \boldsymbol{\alpha}^\top(\tau)} + (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \frac{\partial \text{vec} [\boldsymbol{\omega}(\tau)]}{\partial \boldsymbol{\alpha}^\top(\tau)} \\ &= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_d) \left(\sum_{m=0}^{j-1} \mathbf{J}(\boldsymbol{\Phi}^\top(\tau))^{j-1-m} \otimes \boldsymbol{\Psi}_m(\tau) \right) \\ &\quad + (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \left(\mathbf{N}_1^\top(\tau) \mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau) \mathbf{N}_2(\tau) \right)^{-1} \mathbf{N}_2^\top(\tau) \mathbf{D}_2(\tau). \end{aligned}$$

For $\nabla_{\boldsymbol{\sigma}(\tau)} \mathbf{B}_j(\tau)$,

$$\begin{aligned} \nabla_{\boldsymbol{\sigma}(\tau)} \mathbf{B}_j(\tau) &= \frac{\partial \text{vec} [\boldsymbol{\Psi}_j(\tau) \boldsymbol{\omega}(\tau)]}{\partial \boldsymbol{\sigma}^\top(\tau)} = (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \frac{\partial \text{vec} [\boldsymbol{\omega}(\tau)]}{\partial \boldsymbol{\sigma}^\top(\tau)} \\ &= (\mathbf{I}_d \otimes \boldsymbol{\Psi}_j(\tau)) \mathbf{L}_d^\top \left(\mathbf{N}_1^\top(\tau) \mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau) \mathbf{N}_2(\tau) \right)^{-1} \mathbf{N}_1^\top(\tau) \mathbf{D}_1. \end{aligned}$$

The proof is now completed. □

Proof of Theorem 2.3.

We need to prove that $\lim_{T \rightarrow \infty} \Pr(\text{IC}(\mathbf{p}) < \text{IC}(p)) = 0$ for all $\mathbf{p} \neq p$ and $\mathbf{p} \leq \mathbf{P}$.

Note that

$$\text{IC}(\mathbf{p}) - \text{IC}(p) = \log[\text{RSS}(\mathbf{p})/\text{RSS}(p)] + (\mathbf{p} - p)\chi_T.$$

(i) For $\mathbf{p} < p$, Lemma B.7 implies that $\text{RSS}(\mathbf{p})/\text{RSS}(p) > 1 + \nu$ for some $\nu > 0$ with large probability for all large T . Thus, $\log[\text{RSS}(\mathbf{p})/\text{RSS}(p)] \geq \nu/2$ for large T . Because $\chi_T \rightarrow 0$, we have $\text{IC}(\mathbf{p}) - \text{IC}(p) \geq \nu/2 - (p - \mathbf{p})\chi_T \geq \nu/3$ for large T with large probability. Thus $\Pr(\text{IC}(\mathbf{p}) < \text{IC}(p)) \rightarrow 0$ for $\mathbf{p} < p$.

(ii) We then consider $\mathbf{p} > p$. Lemma B.7 implies that $\log[\text{RSS}(\mathbf{p})/\text{RSS}(p)] = 1 + O_P(\rho_T^2)$ with $\rho_T = h^2 + \sqrt{\frac{\log(T)}{Th}}$. Hence, $\log[\text{RSS}(\mathbf{p})/\text{RSS}(p)] = O_P(\rho_T^2)$. Because $(\mathbf{p} - p)\chi_T \geq \chi_T$, which converges to zero at a slower rate than ρ_T^2 , it follows that

$$\Pr(\text{IC}(\mathbf{p}) < \text{IC}(p)) \leq \Pr(O_P(\rho_T^2) + \chi_T < 0) \rightarrow 0.$$

The proof is now completed. □

Proof of Lemma 3.1.

By part(1) of Theorem 2.1, we have $\sup_{\tau \in [0,1]} \|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\| = O_P(h^2 + \sqrt{\log T/(Th)})$. Hence, by $Th^6 \rightarrow 0$, we have

$$\sqrt{T}(\widehat{\mathbf{c}} - \mathbf{c}) = \sqrt{T} \int_h^{1-h} \mathbf{C} \left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right) d\tau + o_P(1).$$

In addition, by the proof of part(1) of Theorem 2.1, the uniform convergence results of Lemmas B.5-B.6 and the condition that $\boldsymbol{\beta}(\tau)$ has third-order derivative, we have

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) &= \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\beta}^{(2)}(\tau) + \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau) \left(\frac{1}{Th} \sum_{t=1}^T \mathbf{Z}_{t-1}\boldsymbol{\eta}_t K\left(\frac{\tau_t - \tau}{h}\right) \right) \\ &\quad + O_P(h^2\sqrt{\log T/(Th)}) + O_P(h^3) + O_P(\log T/(Th)) \end{aligned}$$

uniformly over $\tau \in [h, 1-h]$.

As $Th^6 \rightarrow 0$ and $Th^2/(\log T)^2 \rightarrow \infty$, we have

$$\begin{aligned} &\sqrt{T} \left(\widehat{\mathbf{c}} - \mathbf{c} - \frac{1}{2}h^2\widetilde{c}_2 \int_0^1 \mathbf{C}\boldsymbol{\beta}^{(2)}(\tau)d\tau \right) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{C} \left(\int_{-1}^1 \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau) K\left(\frac{\tau_t - \tau}{h}\right) d\tau \right) \mathbf{Z}_{t-1}\boldsymbol{\eta}_t + o_P(1) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau_t) \mathbf{Z}_{t-1}\boldsymbol{\eta}_t + o_P(1) \\ &\rightarrow_D N \left(\mathbf{0}, \int_0^1 \mathbf{C} (\boldsymbol{\Sigma}^{-1}(\tau) \otimes \boldsymbol{\Omega}(\tau)) \mathbf{C}^\top d\tau \right) \end{aligned}$$

by Lemma B.1 under the conditions of this lemma.

The proof is now completed. □

Proof of Theorem 3.1.

First, we introduce some additional notation to facilitate the development. Let $A_Q = \widetilde{v}_0 \cdot \text{tr} \left\{ \int_0^1 \boldsymbol{\Sigma}_Q(\tau) d\tau \right\}$ and $B_Q = 4C_B \cdot \text{tr} \left\{ \int_0^1 \boldsymbol{\Sigma}_Q(\tau)^2 d\tau \right\}$, where $\boldsymbol{\Sigma}_Q(\tau) = \mathbf{H}(\tau)^{1/2} \mathbf{C} \mathbf{V}_\beta(\tau) \mathbf{C}^\top \mathbf{H}(\tau)^{1/2}$. Recall $\rho_T = h^2 + \sqrt{\frac{\log T}{Th}}$. By Theorem 2.1.1, we have

$$\sup_{\tau \in [0,1]} \left\| \widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| = O_P(\rho_T). \quad (\text{A.2})$$

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

$$\int_{\mathcal{B}_T} \left[\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \mathbf{c} \right]^\top \mathbf{H}(\tau) \left[\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \mathbf{c} \right] d\tau = O_P(\log T/T + h^5) = o_P\left(T^{-1}h^{-1/2}\right), \quad (\text{A.3})$$

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

In addition, by Lemma B.5 and Lemma B.6.1, we have

$$\begin{aligned} \sup_{[h,1-h]} \|\mathbf{S}_T(\tau) - \boldsymbol{\Sigma}_{\mathbf{Z}}(\tau) \otimes \boldsymbol{\Lambda}_1\| &= O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right), \\ \sup_{[0,1]} \|\mathbf{R}_T(\tau)\| &= O_P\left(\sqrt{\frac{\log T}{Th}}\right), \end{aligned} \quad (\text{A.4})$$

where $\Sigma_{\mathbf{Z}}(\tau) = \Sigma(\tau) \otimes \mathbf{I}_d$, $\Lambda_1 = \text{diag}(\tilde{c}_0, \tilde{c}_2)$ and $\mathbf{R}_T(\tau) = \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{t-1} \boldsymbol{\eta}_t K_h(\tau_t - \tau)$. Hence, under the null hypothesis, we have

$$\sup_{[h, 1-h]} \left\| \mathbf{C} \hat{\boldsymbol{\beta}}(\tau) - \mathbf{c} - \mathbf{C} \Sigma_{\mathbf{Z}}^{-1}(\tau) \mathbf{R}_T(\tau) \right\| = O_P(\rho_T^2). \quad (\text{A.5})$$

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

$$\begin{aligned} & \int_0^1 \left[\mathbf{C} \hat{\boldsymbol{\beta}}(\tau) - \mathbf{c} \right]^\top \mathbf{H}(\tau) \left[\mathbf{C} \hat{\boldsymbol{\beta}}(\tau) - \mathbf{c} \right] d\tau \\ &= \int_h^{1-h} \mathbf{R}_T^\top(\tau) \mathbf{H}_0(\tau) \mathbf{R}_T(\tau) d\tau + O_P(\log T/T + h^5) + O_P\left(\rho_T^2 \sqrt{\frac{\log T}{Th}}\right) \\ &= \int_0^1 \mathbf{R}_T^\top(\tau) \mathbf{H}_0(\tau) \mathbf{R}_T(\tau) d\tau + o_P(T^{-1}h^{-1/2}), \end{aligned}$$

where $\mathbf{H}_0(\tau) = \Sigma_{\mathbf{Z}}^{-1}(\tau) \mathbf{C}^\top \mathbf{H}(\tau) \mathbf{C} \Sigma_{\mathbf{Z}}^{-1}(\tau)$.

Consider $\int_0^1 \mathbf{R}_T^\top(\tau) \mathbf{H}_0(\tau) \mathbf{R}_T(\tau) d\tau$, and write

$$\begin{aligned} & \int_0^1 \mathbf{R}_T^\top(\tau) \mathbf{H}_0(\tau) \mathbf{R}_T(\tau) d\tau = \frac{1}{T^2 h^2} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \left\{ \int_0^1 \mathbf{H}_0(\tau) K^2\left(\frac{\tau - \tau_t}{h}\right) d\tau \right\} \mathbf{Z}_{t-1} \boldsymbol{\eta}_t \\ &+ \frac{1}{T^2 h^2} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \left\{ \int_0^1 \mathbf{H}_0(\tau) K\left(\frac{\tau - \tau_t}{h}\right) K\left(\frac{\tau - \tau_s}{h}\right) d\tau \right\} \mathbf{Z}_{s-1} \boldsymbol{\eta}_s \\ &:= I_{T,1} + I_{T,2}, \end{aligned}$$

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

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

$$\begin{aligned} & \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \Sigma_{\mathbf{Z}}^{-1}(\tau) \mathbf{C}^\top \mathbf{H}(\tau) \mathbf{C} \Sigma_{\mathbf{Z}}^{-1}(\tau) \mathbf{Z}_{t-1} \boldsymbol{\eta}_t \\ &= \text{tr} \left\{ \left[\left(\Sigma^{-1}(\tau) \mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top \Sigma^{-1}(\tau) \right) \otimes \boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top \right] \cdot \mathbf{C}^\top \mathbf{H}(\tau) \mathbf{C} \right\}. \end{aligned}$$

Then we have

$$\begin{aligned} I_{T,1} &= \tilde{v}_0 \frac{1}{T^2 h} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \left[\mathbf{H}_0(\tau_t) + O(h) \right] \mathbf{Z}_{t-1} \boldsymbol{\eta}_t \\ &= \tilde{v}_0 \frac{1}{T^2 h} \sum_{t=1}^T \text{tr} \left\{ \left[\left(\Sigma^{-1}(\tau_t) \mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top \Sigma^{-1}(\tau_t) \right) \otimes \boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top \right] \cdot \mathbf{C}^\top \mathbf{H}(\tau_t) \mathbf{C} \right\} + O_P(T^{-1}) \\ &= \tilde{v}_0 \frac{1}{T^2 h} \sum_{t=1}^T \text{tr} \left\{ \left[\Sigma^{-1}(\tau_t) \otimes \boldsymbol{\Omega}(\tau_t) \right] \cdot \mathbf{C}^\top \mathbf{H}(\tau_t) \mathbf{C} \right\} + O_P(T^{-1} + T^{-3/2} h^{-1}) \\ &\rightarrow_P (Th)^{-1} A_Q. \end{aligned}$$

Consider $I_{T,2}$. Let $w_{s,t} = \frac{1}{T\sqrt{h}} \int_{-1}^1 K(u) K(u + \frac{t-s}{Th}) du$. Since

$$\int_0^1 \mathbf{H}_0(\tau) K\left(\frac{\tau - \tau_t}{h}\right) K\left(\frac{\tau - \tau_s}{h}\right) d\tau = h \int_{-1}^1 \mathbf{H}_0(\tau_t + uh) K(u) K\left(u + \frac{t-s}{Th}\right) du,$$

we have

$$T\sqrt{h}I_{T,2} = 2 \sum_{t=2}^T \sum_{s=1}^{t-1} \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \mathbf{H}_0(\tau_t) \mathbf{Z}_{s-1} \boldsymbol{\eta}_s w_{s,t} (1 + o(1)) = 2\tilde{U} + o_P(1),$$

where the definition of \tilde{U} is obvious. By Lemma B.12, we have

$$\tilde{U} \rightarrow_D N\left(0, \sigma_{\tilde{U}}^2\right),$$

where $\sigma_{\tilde{U}}^2 = \lim_{T \rightarrow \infty} \sum_{t=2}^T \text{tr} \left\{ E \left(\mathbf{H}_0^\top(\tau_t) \mathbf{Z}_{t-1} \boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \mathbf{H}_0(\tau_t) \right) E \left(\sum_{s=1}^{t-1} \mathbf{Z}_{s-1} \boldsymbol{\eta}_s \boldsymbol{\eta}_s^\top \mathbf{Z}_{s-1}^\top \right) w_{s,t}^2 \right\}$.

We then show that $\sigma_{\tilde{U}}^2 = C_B \text{tr} \left\{ \int_0^1 \boldsymbol{\Sigma}_Q(\tau)^2 d\tau \right\}$. Let $\mathbf{V}_1(\tau) = \mathbf{H}_0(\tau) \mathbf{V}_2(\tau) \mathbf{H}_0(\tau)$ and $\mathbf{V}_2(\tau) = \boldsymbol{\Sigma}(\tau) \otimes \boldsymbol{\Omega}(\tau)$. Write

$$\begin{aligned} & \sum_{t=2}^T E \left(\mathbf{H}_0^\top(\tau_t) \mathbf{Z}_{t-1} \boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top \mathbf{H}_0(\tau_t) \right) E \left(\sum_{s=1}^{t-1} \mathbf{Z}_{s-1} \boldsymbol{\eta}_s \boldsymbol{\eta}_s^\top \mathbf{Z}_{s-1}^\top \right) w_{s,t}^2 \\ &= \sum_{t=2}^T \sum_{s=1}^{t-1} \mathbf{V}_1(\tau_t) \mathbf{V}_2(\tau_s) w_{s,t}^2 = \frac{1}{T^2 h} \sum_{t=2}^T \sum_{s=1}^{t-1} \mathbf{V}_1(\tau_t) \mathbf{V}_2(\tau_s) \left[\int_{-1}^1 K(u) K\left(u + \frac{t-s}{Th}\right) du \right]^2 \\ &= \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} \mathbf{V}_1(\tau_s + j/T) \mathbf{V}_2(\tau_s) \left[\int_{-1}^1 K(u) K\left(u + \frac{t-s}{Th}\right) du \right]^2 \\ &= \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} \mathbf{V}_1(\tau_s + j/T) \mathbf{V}_2(\tau_s) \left[\int_{-1}^1 K(u) K\left(u + \frac{j}{Th}\right) du \right]^2 \\ &= \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} \mathbf{V}_1(\tau_s) \mathbf{V}_2(\tau_s) \left[\int_{-1}^1 K(u) K\left(u + \frac{j}{Th}\right) du \right]^2 \\ & \quad + \frac{1}{T^2 h} \sum_{s=1}^{T-1} \sum_{j=1}^{T-s} O(j/T) \mathbf{V}_2(\tau_s) \left[\int_{-1}^1 K(u) K\left(u + \frac{j}{Th}\right) du \right]^2 := I_{T,3} + I_{T,4}, \end{aligned}$$

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

It is easy to verify $\text{tr} \{I_{T,3}\} \rightarrow C_B \text{tr} \left\{ \int_0^1 \boldsymbol{\Sigma}_Q(\tau)^2 d\tau \right\}$. For $I_{T,4}$, we have

$$\begin{aligned} \|I_{T,4}\| &\leq M \frac{1}{Th} \sum_{j=1}^T j/T \left[\int_{-1}^1 K(u) K\left(u + \frac{j}{Th}\right) du \right]^2 \\ &\simeq M \int_0^2 v h \left[\int_{-1}^1 K(u) K(u+v) du \right]^2 dv = o(1). \end{aligned}$$

Combining the above results, we have proved

$$T\sqrt{h} \left[\int_0^1 \mathbf{R}_T^\top(\tau) \mathbf{H}_0(\tau) \mathbf{R}_T(\tau) d\tau - (Th)^{-1} A_Q \right] \rightarrow_D N(0, B_Q).$$

Note that

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

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

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

$$\begin{aligned} I_{T,6} &= (\widehat{\mathbf{c}} - \mathbf{c})^\top \int_h^{1-h} \mathbf{H}(\tau) \mathbf{C} \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau) \mathbf{R}_T(\tau) d\tau + o_P(T^{-1}h^{-1/2}) \\ &= O_P(T^{-1}) + o_P(T^{-1}h^{-1/2}) = o_P(T^{-1}h^{-1/2}) \end{aligned}$$

provided that

$$\begin{aligned} &\int_h^{1-h} \mathbf{H}(\tau) \mathbf{C} \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau) \mathbf{R}_T(\tau) d\tau \\ &= \frac{1}{T} \sum_{t=1}^T \int_{-1}^1 \mathbf{H}(\tau_t + uh) \mathbf{C} \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau_t + uh) K(u) du \mathbf{Z}_{t-1} \boldsymbol{\eta}_t = O_P(T^{-1/2}). \end{aligned}$$

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

Observe that

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

Then we just need to focus on

$$\int_0^1 [\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\mathbf{c}}]^\top \widehat{\mathbf{H}}(\tau) [\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\mathbf{c}}] d\tau - \int_0^1 [\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\mathbf{c}}]^\top \mathbf{H}(\tau) [\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\mathbf{c}}] d\tau := I_{T,7}$$

Hence, it suffices to show $T\sqrt{h}I_{T,7} = o_P(1)$. Using Lemma B.6.3, it is easy to know that

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

The proof is now completed. □

Proof of Corollary 3.1.

Under the local alternative (3.7), we have $\mathbf{C}\boldsymbol{\beta}(\tau) = \mathbf{c} + d_T \mathbf{f}(\tau)$ and thus

$$\begin{aligned} &\widehat{\mathbf{Q}}_{\mathbf{C},\mathbf{H}} - \int_0^1 \mathbf{R}_T^\top(\tau) \mathbf{H}_0(\tau) \mathbf{R}_T(\tau) d\tau \\ &= d_T^2 \int_0^1 \mathbf{f}(\tau)^\top \mathbf{H}(\tau) \mathbf{f}(\tau) d\tau + 2d_T \int_0^1 \mathbf{f}(\tau)^\top \mathbf{H}(\tau) (\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \mathbf{C}\boldsymbol{\beta}(\tau)) d\tau \\ &\quad + \left[\int_0^1 (\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \mathbf{C}\boldsymbol{\beta}(\tau))^\top \mathbf{H}(\tau) (\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \mathbf{C}\boldsymbol{\beta}(\tau)) d\tau - \int_0^1 \mathbf{R}_T^\top(\tau) \mathbf{H}_0(\tau) \mathbf{R}_T(\tau) d\tau \right] \\ &= d_T^2 \int_0^1 \mathbf{f}(\tau)^\top \mathbf{H}(\tau) \mathbf{f}(\tau) d\tau + I_{T,1} + I_{T,2}. \end{aligned}$$

Since $\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \mathbf{C}\boldsymbol{\beta}(\tau) = O_P\left(d_T\rho_T + \sqrt{\frac{\log T}{Th}}\rho_T\right) + \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau)\mathbf{R}_T(\tau)$ uniformly over $\tau \in [h, 1-h]$ and

$$\int_0^1 \mathbf{f}(\tau)^\top \mathbf{H}(\tau) \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau) \mathbf{R}_T(\tau) d\tau = O_P(T^{-1/2}),$$

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 $\sup_{\tau \in [0,1]} \|\mathbf{R}_T(\tau)\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right)$, 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 $T\sqrt{h}\left(\int_0^1 \mathbf{R}_T^\top(\tau)\mathbf{H}_0(\tau)\mathbf{R}_T(\tau) - (Th)^{-1}A_Q\right) \rightarrow N(0, B_Q)$, we have

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

In addition, similar to the proofs of Theorem 3.1, we have $T\sqrt{h}\left(\widehat{Q}_{\mathbf{C},\mathbf{H}} - \widehat{Q}_{\mathbf{C},\widehat{\mathbf{H}}}\right) = o_P(1)$. The proof is now completed. \square

Appendix B

In this appendix, we present the results omitted from the main text of this paper. Specifically, Appendix B.1 includes the preliminary lemmas. Appendix B.2 provides the proofs of the preliminary lemmas.

B.1 Preliminary Lemmas

Lemma B.1. *Suppose $\{Z_t, \mathcal{F}_t\}$ is a martingale difference sequence, $S_T = \sum_{t=1}^T Z_t$, $U_T = \sum_{t=1}^T Z_t^2$ and $s_T^2 = E(U_T^2) = E(S_T^2)$. If $s_T^{-2} U_T^2 \rightarrow_P 1$ and $\sum_{t=1}^T E[Z_{T,t}^2 I(|Z_{T,t}| > \nu)] \rightarrow 0$ for any $\nu > 0$ with $Z_{T,t} = s_T^{-1} Z_t$, then as $T \rightarrow \infty$, $s_T^{-1} S_T \rightarrow_D N(0, 1)$.*

Lemma B.1 can be found in Hall and Heyde (1980).

Lemma B.2. *Let $\{Z_t, \mathcal{F}_t\}$ be a martingale difference sequence. Suppose that $|Z_t| \leq M$ for a constant M , $t = 1, \dots, T$. Let $V_T = \sum_{t=1}^T \text{Var}(Z_t | \mathcal{F}_{t-1}) \leq V$ for some $V > 0$. Then for any given $\nu > 0$,*

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

Lemma B.2 is the Proposition 2.1 of Freedman (1975).

Assumption A.1. $\max_t \|\boldsymbol{\mu}_t\| < \infty$, $\max_t \sum_{j=1}^{\infty} j \|\mathbf{B}_{j,t}\| < \infty$, $\limsup_{T \rightarrow \infty} \sum_{t=1}^{T-1} \|\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\| < \infty$ and $\limsup_{T \rightarrow \infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \|\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}\| < \infty$. Define the stochastic process of the form $\mathbf{h}_t = \boldsymbol{\mu}_t + \sum_{j=0}^{\infty} \mathbf{B}_{j,t} \boldsymbol{\epsilon}_t$ for $t = 1, \dots, T$.

Lemma B.3. *Let Assumptions 2 and A.1 hold and $\max_{t \geq 1} E \left(\|\boldsymbol{\epsilon}_t\|^\delta | \mathcal{F}_{t-1} \right) < \infty$ a.s.. In addition, let $\{\mathbf{W}_{T,t}(\cdot)\}_{t=1}^T$ be a sequence of $q \times d$ 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. $\sup_{\tau \in [a, b]} \sum_{t=1}^T \|\mathbf{W}_{T,t}(\tau)\| = O(1)$;
2. $\sup_{\tau \in [a, b]} \sum_{t=1}^{T-1} \|\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)\| = O(d_T)$, where $d_T = \sup_{\tau \in [a, b], t \geq 1} \|\mathbf{W}_{T,t}(\tau)\|$.

Then as $T \rightarrow \infty$,

1. $\sup_{\tau \in [a, b]} \left\| \sum_{t=1}^T \mathbf{W}_{T,t}(\tau) (\mathbf{h}_t - E(\mathbf{h}_t)) \right\| = O_P(\sqrt{d_T \log T})$ provided $T^{\frac{2}{3}} d_T \log T \rightarrow 0$;
2. $\sup_{\tau \in [a, b]} \left\| \sum_{t=1}^T \mathbf{W}_{T,t}(\tau) (\mathbf{h}_t \mathbf{h}_{t+p}^\top - E(\mathbf{h}_t \mathbf{h}_{t+p}^\top)) \right\| = O_P(\sqrt{d_T \log T})$ for any fixed integer $p \geq 0$ provided $T^{\frac{4}{3}} d_T \log T \rightarrow 0$, where δ is the same as in Assumption 2.

Lemma B.3 is the Theorem 2.1 in Yan et al. (2020).

Lemma B.4. *Let Assumptions 2 and A.1 hold. In addition, let $\{\mathbf{W}_{T,t}(\cdot)\}_{t=1}^T$ be a sequence of $q \times d$ 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. $\sup_{\tau \in [a, b]} \sum_{t=1}^T \|\mathbf{W}_{T,t}(\tau)\| = O(1)$;
2. $\sup_{\tau \in [a, b]} \sum_{t=1}^{T-1} \|\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)\| = O(d_T)$, where $d_T = \sup_{\tau \in [a, b], t \geq 1} \|\mathbf{W}_{T,t}(\tau)\|$.

Then as $T \rightarrow \infty$, for any $\tau \in [a, b]$

1. $\|\sum_{t=1}^T \mathbf{W}_{T,t}(\tau) (\mathbf{h}_t - E(\mathbf{h}_t))\| = O_P(\sqrt{d_T})$;
2. $\|\sum_{t=1}^T \mathbf{W}_{T,t}(\tau) (\mathbf{h}_t \mathbf{h}_{t+p}^\top - E(\mathbf{h}_t \mathbf{h}_{t+p}^\top))\| = O_P(\sqrt{d_T})$ for any fixed integer $p \geq 0$.

The proof of Lemma B.4 is similar with that of Theorem 2.1 in Yan et al. (2020).

Lemma B.5. Suppose Assumptions 1-3 hold. Let $\mathbf{W}(\cdot)$ be a twice-differentiable functional matrix in $\mathbb{R}^{m \times d}$. As $T \rightarrow \infty$,

1. for $\tau \in (0, 1)$,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \mathbf{W}(\tau_t) \mathbf{x}_t \left(\frac{\tau_t - \tau}{h} \right)^k K_h(\tau_t - \tau) - \tilde{c}_k \mathbf{W}(\tau) \boldsymbol{\mu}(\tau) &= O_P(h^2 + 1/(\sqrt{Th})), \\ \frac{1}{T} \sum_{t=1}^T \mathbf{W}(\tau_t) \mathbf{x}_t \mathbf{x}_{t+p}^\top \left(\frac{\tau_t - \tau}{h} \right)^k K_h(\tau_t - \tau) - \tilde{c}_k \mathbf{W}(\tau) \boldsymbol{\Sigma}_p(\tau) &= O_P(h^2 + 1/(\sqrt{Th})), \end{aligned}$$

where $\boldsymbol{\Sigma}_p(\tau) = \boldsymbol{\mu}(\tau) \boldsymbol{\mu}^\top(\tau) + \sum_{j=0}^{\infty} \mathbf{B}_j(\tau) \mathbf{B}_{j+p}^\top(\tau)$ for fixed integers k and $p \geq 0$;

2. given $\frac{T^{1-\frac{2}{\delta}} h}{\log T} \rightarrow \infty$ and $\rho_T = h^2 + \sqrt{\frac{\log(T)}{Th}}$,

$$\sup_{\tau \in [h, 1-h]} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{W}(\tau_t) \mathbf{x}_t \left(\frac{\tau_t - \tau}{h} \right)^k K_h(\tau_t - \tau) - \tilde{c}_k \mathbf{W}(\tau) \boldsymbol{\mu}(\tau) \right\| = O_P(\rho_T^2);$$

3. given $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \rightarrow \infty$ and $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s.,

$$\sup_{\tau \in [h, 1-h]} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{W}(\tau_t) \mathbf{x}_t \mathbf{x}_{t+p}^\top \left(\frac{\tau_t - \tau}{h} \right)^k K_h(\tau_t - \tau) - \tilde{c}_k \mathbf{W}(\tau) \boldsymbol{\Sigma}_p(\tau) \right\| = O_P(\rho_T^2).$$

4. $\frac{1}{T} \sum_{t=1}^T \mathbf{W}(\tau_t) (\mathbf{x}_t - E(\mathbf{x}_t)) = O_P(1/\sqrt{T})$ and $\frac{1}{T} \sum_{t=1}^T \mathbf{W}(\tau_t) (\mathbf{x}_t \mathbf{x}_{t+p}^\top - E(\mathbf{x}_t \mathbf{x}_{t+p}^\top)) = O_P(1/\sqrt{T})$.

Lemma B.6. Let Assumptions 1-3 hold. Suppose $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \rightarrow \infty$ and $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s. As $T \rightarrow \infty$,

1. $\sup_{\tau \in [0, 1]} \left\| \frac{1}{Th} \sum_{t=1}^T \mathbf{Z}_{t-1} \boldsymbol{\eta}_t K\left(\frac{\tau_t - \tau}{h}\right) \right\| = O_P\left(\left(\frac{\log T}{Th}\right)^{\frac{1}{2}}\right)$;
2. $\frac{1}{\sqrt{Th}} \sum_{t=1}^T \boldsymbol{\eta}_t (\boldsymbol{\eta}_t - \hat{\boldsymbol{\eta}}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) = o_P(1)$ for $\forall \tau \in [0, 1]$;
3. $\sup_{\tau \in [h, 1-h]} \left\| \hat{\mathbf{V}}_\beta(\tau) - \mathbf{V}_\beta(\tau) \right\| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{\frac{1}{2}}\right)$.

Lemma B.7. Let Assumptions 1-3 hold. Suppose $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \rightarrow \infty$ and $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s. As $T \rightarrow \infty$,

1. if $\mathbf{p} \geq p$, then $RSS(\mathbf{p}) = \frac{1}{T} \sum_{t=1}^T E(\boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t) + O_P(\rho_T^2)$ with $\rho_T = h^2 + \sqrt{\frac{\log(T)}{Th}}$;
2. if $\mathbf{p} < p$, then $RSS(\mathbf{p}) = \frac{1}{T} \sum_{t=1}^T E(\boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t) + c + o_P(1)$ with some constant $c > 0$.

Define $w_{s,t} = \frac{1}{T\sqrt{h}} \int_{-1}^1 K(u) K(u + \frac{t-s}{Th}) du$. Let $a_t = \sum_{s=1}^{t-1} w_{s,t}^2$, $b_s = \sum_{t=s+1}^T w_{s,t}^2$ and $\sigma^2 = \sum_{t=2}^T a_t$.

Lemma B.8. *Suppose Assumption 3 hold. Then, we have*

1. $\sigma^2 \rightarrow \int_0^2 \left[\int_{-1}^{1-v} K(u) K(u+v) du \right]^2 dv$;
2. $\max_{2 \leq t \leq T} a_t = O(1/T)$;
3. *for any fixed* $J \in \mathbb{N}$, $\sum_{s=1}^{T-J} w_{s,s+J}^2 = O(1/(Th))$;
4. $T \sum_{s=1}^{T-1} b_s^2 = O(1)$;
5. $\sum_{k=1}^{T-1} \sum_{t=1}^{k-1} \left[\sum_{j=k+1}^T w_{k,j} w_{t,j} \right]^2 = O(1/T)$;

For a random vector \mathbf{z} , we write $\mathbf{z} \in \mathcal{L}^q$, $q > 0$, if $\|\mathbf{z}\|_q = [E(\|\mathbf{z}\|^q)]^{1/q} < \infty$, and we denote $\|\cdot\| = \|\cdot\|_2$.

Lemma B.9. *Let $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T$ be a d -dimensional martingale difference for which $\boldsymbol{\xi}_t \in \mathcal{L}^p$, $p > 1$. Let $p^* = \min(2, p)$. Then*

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

Let $\mathbf{h}_{t-1}^* := \sum_{s=1}^{t-1} w_{s,t} \mathbf{y}_s$, where $\mathbf{y}_t \in \mathcal{L}^\delta$, $t = 1, 2, \dots, T$ are martingale differences subject to the filtration \mathcal{F}_t and $\delta > 4$.

Lemma B.10. *Suppose Assumption 3 hold. Assume $\mathbf{w}_t^* \in \mathcal{L}^{\delta/2}$ for some $\delta > 4$ is \mathcal{F}_t -measurable. Then, as $T \rightarrow \infty$,*

$$E \left[\sum_{t=2}^T \text{tr} \left[(\mathbf{w}_t^* - E(\mathbf{w}_t^*)) \mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top} \right] \right] \rightarrow 0.$$

Lemma B.11. *Suppose Assumptions 1-3 hold and $\mathbf{y}_t = \mathbf{Z}_{t-1} \boldsymbol{\eta}_t$. Then, as $T \rightarrow \infty$,*

$$\sum_{t=2}^T \text{tr} \left[\mathbf{H}_t \left(\mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top} - E(\mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top}) \right) \right] \rightarrow_P 0,$$

where \mathbf{H}_t is a $d^2 \times d^2$ deterministic weighting function satisfying $\|\mathbf{H}_t\| < \infty$.

Let $Q_T = \sum_{t=2}^T \mathbf{y}_t^\top \mathbf{H}_t \mathbf{h}_{t-1}^*$, where \mathbf{H}_t is a $d^2 \times d^2$ functional weighting matrix.

Lemma B.12. *Suppose Assumptions 1-3 hold and $\mathbf{y}_t = \mathbf{Z}_{t-1} \boldsymbol{\eta}_t$. Then as $T \rightarrow \infty$,*

$$Q_T \rightarrow_D N(0, \sigma_Q^2).$$

where $\sigma_Q^2 = \lim_{T \rightarrow \infty} \sum_{t=2}^T \text{tr} \left[E(\mathbf{H}_t^\top \mathbf{y}_t \mathbf{y}_t^\top \mathbf{H}_t) E(\mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top}) \right]$.

B.2 Proofs of the Preliminary Lemmas

Proof of Lemma B.5.

(1). First, for any fixed $\tau \in (0, 1)$, let $\mathbf{W}_{T,t}(\tau) = \frac{1}{T} \mathbf{W}(\tau_t) \left(\frac{\tau_t - \tau}{h} \right)^k K_h(\tau_t - \tau)$. It is straightforward to show that $\sup_{\tau \in [0,1]} \sum_{t=1}^T \|\mathbf{W}_{T,t}(\tau)\| = O(1)$, $\sup_{\tau \in [0,1], t \geq 1} \|\mathbf{W}_{T,t}(\tau)\| = O(1/(Th))$ and

$$\sup_{\tau \in [0,1]} \sum_{t=1}^{T-1} \|\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)\| = O(1/(Th)).$$

Second, by triangle inequality, Cauchy-Schwarz inequality and Proposition 2.1,

$$\begin{aligned} \sum_{t=1}^T \left\| \mathbf{x}_t \mathbf{x}_t^\top - \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^\top \right\| &\leq \sum_{t=1}^T (\|\mathbf{x}_t\| + \|\tilde{\mathbf{x}}_t\|) \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\| \\ &\leq \left(\sum_{t=1}^T (\|\mathbf{x}_t\| + \|\tilde{\mathbf{x}}_t\|)^2 \right)^{1/2} \left(\sum_{t=1}^T \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|^2 \right)^{1/2} = O_P(\sqrt{T}) \cdot O_P(1/\sqrt{T}) = O_P(1). \end{aligned}$$

In addition, similar to the proof of Proposition 2.1, we have

$$\sum_{j=0}^{\infty} j \|\mathbf{B}_j(\tau)\| = \sum_{j=0}^{\infty} j \|\Psi_j(\tau) \boldsymbol{\omega}(\tau)\| \leq M \sum_{j=0}^{\infty} j \rho_A^j < \infty,$$

and

$$\begin{aligned} &\sum_{t=1}^{T-1} \sum_{j=0}^{\infty} j \|\mathbf{B}_j(\tau_{t+1}) - \mathbf{B}_j(\tau_t)\| \\ &\leq \sup_{\tau \in [0,1]} \sum_{j=1}^{\infty} j \|\Psi_j(\tau)\| \sum_{t=1}^T \|\boldsymbol{\omega}(\tau_{t+1}) - \boldsymbol{\omega}(\tau_t)\| + M \sum_{t=1}^{T-1} \sum_{j=0}^{\infty} j \|\Psi_j(\tau_{t+1}) - \Psi_j(\tau_t)\| \\ &\leq M \sum_{j=0}^{\infty} (j^2 \rho_A^{j-1} + j \rho_A^j) < \infty. \end{aligned}$$

Therefore, Lemma B.4 are still valid for the TV-VAR(p) process. The proof of part (1) is completed.

(2)-(4). The proofs of part (2)-(4) can be done in a similar way to that of part (1). \square

Proof of Lemma B.6.

(1). Let $\{S_l\}$ be a finite number of sub-intervals covering 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,

$$\begin{aligned} \sup_{\tau \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{t-1} \boldsymbol{\eta}_t K_h(\tau_t - \tau) \right\| &\leq \max_{1 \leq l \leq N_T} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{t-1} \boldsymbol{\eta}_t K_h(\tau_t - s_l) \right\| \\ &\quad + \max_{1 \leq l \leq N_T} \sup_{\tau \in S_l} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{t-1} \boldsymbol{\eta}_t (K_h(\tau_t - \tau) - K_h(\tau_t - s_l)) \right\| \\ &:= I_{T,1} + I_{T,2}. \end{aligned}$$

By the continuity of kernel function $K(\cdot)$ and taking $\delta_T = O(\gamma_T h^2)$ with $\gamma_T = \left(\frac{\log T}{Th}\right)^{\frac{1}{2}}$, then we have

$$E|I_{T,2}| \leq M \frac{\delta_T}{h^2} E \|\mathbf{Z}_{t-1} \boldsymbol{\eta}_t\| = O(\gamma_T).$$

We then apply the truncation method again. Define $\mathbf{u}_t = \mathbf{Z}_{t-1} \boldsymbol{\eta}_t$, $\mathbf{u}'_t = \mathbf{u}_t I(\|\mathbf{u}_t\| \leq T^{\frac{2}{\delta}})$ and $\mathbf{u}''_t = \mathbf{u}_t - \mathbf{u}'_t$. Then we have

$$I_{T,1} = \max_{1 \leq l \leq N_T} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{u}'_t + \mathbf{u}''_t - E(\mathbf{u}'_t + \mathbf{u}''_t | \mathcal{F}_{t-1}) K_h(\tau_t - s_l) \right\|$$

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

Now consider $I_{T,12}$. Let $d_T = \max_{1 \leq t \leq T, 1 \leq l \leq N_T} K_h(\tau_t - s_l)/T$. By Holder's inequality and Chebyshev inequality, we have

$$\begin{aligned}
E|I_{T,12}| &\leq d_T \sum_{t=1}^T E \|\boldsymbol{\eta}''_t\| \leq d_T \sum_{t=1}^T E \left(\|\mathbf{Z}_{t-1} \boldsymbol{\eta}_t\|^{\delta/2} \right)^{\frac{2}{\delta}} \left(\frac{E \left(\|\mathbf{Z}_{t-1} \boldsymbol{\eta}_t\|^{\delta/2} \right)}{T} \right)^{\frac{\delta-2}{\delta}} \\
&= O(T^{\frac{2}{\delta}} d_T) = o \left(\sqrt{\frac{\log T}{Th}} \right).
\end{aligned}$$

Similarly, $I_{T,13} = O_P(T^{\frac{2}{\delta}} d_T)$.

For any fixed $1 \leq l \leq N_T$, let $\mathbf{Y}_t := \frac{1}{T} (\mathbf{u}'_t - E(\mathbf{u}'_t | \mathcal{F}_{t-1})) K_h(\tau_t - s_l)$, then we have $E(\mathbf{Y}_t | \mathcal{F}_{t-1}) = 0$ and $\|\mathbf{Y}_t\| \leq 2T^{2/\delta} d_T$ with $d_T = \max_{1 \leq t \leq T} K_h(\tau_t - s_l)/T$. Also, by Lemma B.5, we have

$$\sup_{0 \leq \tau \leq 1} \frac{1}{T} \sum_{t=1}^T E \left(\|\mathbf{Z}_{t-1} \boldsymbol{\eta}_t\|^2 | \mathcal{F}_{t-1} \right) K_h(\tau_t - \tau) = O_P(1)$$

which follows that

$$\max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^T E(\mathbf{Y}_t \mathbf{Y}_t^\top | \mathcal{F}_{t-1}) \right\| \leq 4 \max_{1 \leq l \leq N_T} \sum_{t=1}^T E \left(\|\mathbf{u}_t\|^2 | \mathcal{F}_{t-1} \right) \frac{K_h^2(\tau_t - s_l)}{T^2} = O_P(d_T).$$

Therefore, we have $\max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^T E(\mathbf{Y}_t \mathbf{Y}_t^\top | \mathcal{F}_{t-1}) \right\| \leq \frac{M}{Th}$ in probability. By Lemma B.2, we have

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

if $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \rightarrow \infty$, which completes the proof of part (1).

(2). Note that

$$\begin{aligned}
&\frac{1}{\sqrt{Th}} \sum_{t=1}^T \boldsymbol{\eta}_t (\boldsymbol{\eta}_t - \hat{\boldsymbol{\eta}}_t)^\top K \left(\frac{\tau_t - \tau}{h} \right) = \frac{1}{\sqrt{Th}} \sum_{t=1}^T \boldsymbol{\eta}_t \mathbf{z}_{t-1}^\top \left(\hat{\mathbf{A}}(\tau_t) - \mathbf{A}(\tau_t) \right)^\top K \left(\frac{\tau_t - \tau}{h} \right) \\
&= \frac{1}{\sqrt{Th}} \sum_{t=1}^T \boldsymbol{\eta}_t [\mathbf{z}_{t-1}^\top, \mathbf{0}_{(d^2 p + d) \times 1}^\top] \frac{1}{2} h^2 \mathbf{S}_T^{-1}(\tau_t) \begin{pmatrix} \mathbf{S}_{T,2}(\tau_t) \\ \mathbf{S}_{T,3}(\tau_t) \end{pmatrix} \mathbf{A}^{(2),\top}(\tau_t) K \left(\frac{\tau_t - \tau}{h} \right)
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\sqrt{Th}} \sum_{t=1}^T \boldsymbol{\eta}_t [\mathbf{z}_{t-1}^\top, \mathbf{0}_{(d^2 p + d) \times 1}^\top] \mathbf{S}_T^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^T (\mathbf{z}_{s-1}^* \mathbf{z}_{s-1}^\top \otimes \mathbf{I}_d) \mathbf{M}^\top(\tau_s) K\left(\frac{\tau_s - \tau_t}{h}\right) \right) K\left(\frac{\tau_t - \tau}{h}\right) \\
& + \frac{1}{\sqrt{Th}} \sum_{t=1}^T \boldsymbol{\eta}_t [\mathbf{z}_{t-1}^\top, \mathbf{0}_{(d^2 p + d) \times 1}^\top] \mathbf{S}_T^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^T \mathbf{Z}_{s-1} \boldsymbol{\eta}_s^\top K\left(\frac{\tau_s - \tau_t}{h}\right) \right) K\left(\frac{\tau_t - \tau}{h}\right) \\
& := J_{T,1} + J_{T,2} + J_{T,3}.
\end{aligned}$$

For $J_{T,1}$ to $J_{T,2}$, using Lemmas B.5.2-3, we can replace the sample covariance matrix with its converged and deterministic value with rate $O_P\left(\sqrt{\frac{\log T}{Th}}\right)$ and hence it's easy to show that $J_{T,1}$ to $J_{T,2}$ are $o_P(1)$.

For $J_{T,3}$, for notational simplicity, we ignore $\mathbf{S}_T^{-1}(\tau_t)$ and hence,

$$\begin{aligned}
J_{T,3} &= \frac{1}{(Th)^{3/2}} \sum_{t=1}^T \boldsymbol{\eta}_t \mathbf{z}_{t-1}^\top \mathbf{z}_{t-1} \boldsymbol{\eta}_t^\top K(0) K\left(\frac{\tau_t - \tau}{h}\right) \\
& + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \mathbf{z}_{t-1}^\top \mathbf{z}_{t+i-1} \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
& + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_{t+i} \mathbf{z}_{t+i-1}^\top \mathbf{z}_{t-1} \boldsymbol{\eta}_t^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
& := J_{T,31} + J_{T,32} + J_{T,33}.
\end{aligned}$$

It's easy to see $J_{T,31} = O_P((Th)^{-1/2})$. For $J_{T,32}$,

$$\begin{aligned}
J_{T,32} &= \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t E\left(\mathbf{z}_{t-1}^\top \mathbf{z}_{t+i-1}\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
& + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\mathbf{z}_{t-1}^\top \mathbf{z}_{t+i-1} - E\left(\mathbf{z}_{t-1}^\top \mathbf{z}_{t+i-1}\right)\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
& := J_{T,321} + J_{T,322}.
\end{aligned}$$

For $J_{T,321}$,

$$\begin{aligned}
E \|J_{T,321}\|^2 &\leq \frac{1}{(Th)^3} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \left\{ E\left(\mathbf{z}_{t-1}^\top \mathbf{z}_{t+i-1}\right) E\left(\boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t\right) E\left(\boldsymbol{\eta}_{t+i}^\top \boldsymbol{\eta}_{t+i}\right) \right\}^2 K^2\left(\frac{i}{Th}\right) K^2\left(\frac{\tau_t - \tau}{h}\right) \\
& = O\left(\frac{1}{Th}\right),
\end{aligned}$$

which then yields that $J_{T,321} = O_P((Th)^{-1/2})$. For $J_{T,322}$,

$$J_{T,322} = \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \sum_{m=1}^p \left(\mathbf{x}_{t-m}^\top \mathbf{x}_{t+i-m} - E\left(\mathbf{x}_{t-m}^\top \mathbf{x}_{t+i-m}\right) \right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right).$$

For notational simplicity, let $p = 1$ and thus

$$J_{T,322} = \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\boldsymbol{\mu}_{t-1}^\top \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j,t+i-1} \boldsymbol{\eta}_{t+i-1-j} \right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right)$$

$$\begin{aligned}
& + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \boldsymbol{\eta}_{t-1-j}^\top \boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\mu}_{t+i-1} \right) \boldsymbol{\eta}_{t+i}^\top K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right) \\
& + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \left(\boldsymbol{\eta}_{t-1-j}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top - E \left(\boldsymbol{\eta}_{t-1-j}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top \right) \right) \right. \\
& \cdot \text{vec} \left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{j+i,t+i-1} \right) \left. \boldsymbol{\eta}_{t+i}^\top K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right) \right. \\
& \left. + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \sum_{m=0, \neq j+i}^{\infty} \left(\boldsymbol{\eta}_{t+i-1-m}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top \right) \text{vec} \left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{m,t+i-1} \right) \right) \right. \\
& \left. \cdot \boldsymbol{\eta}_{t+i}^\top K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right) \right) \\
& := J_{T,3221} + J_{T,3222} + J_{T,3223} + J_{T,3224}.
\end{aligned}$$

For $J_{T,3221}$,

$$\begin{aligned}
& E \|J_{T,3221}\|^2 \\
& \leq \frac{1}{(Th)^3} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} E \left\| \boldsymbol{\eta}_t \left(\boldsymbol{\mu}_{t-1}^\top \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j,t+i-1} \boldsymbol{\eta}_{t+i-1-j} \right) \right\|^2 E \|\boldsymbol{\eta}_{t+i}\|^2 K^2 \left(\frac{i}{Th} \right) K^2 \left(\frac{\tau_t - \tau}{h} \right) \\
& = O \left(\frac{1}{Th} \right).
\end{aligned}$$

Similarly, $J_{T,3222}$ and $J_{T,3223}$ are $O_P((Th)^{-1/2})$. For $J_{T,3224}$,

$$\begin{aligned}
J_{T,3224} & = \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \left(\boldsymbol{\eta}_t^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top \right) \text{vec} \left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{i-1,t+i-1} \right) \right) \boldsymbol{\eta}_{t+i}^\top K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right) \\
& + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \sum_{m=0, \neq j+i, \neq i-1}^{\infty} \left(\boldsymbol{\eta}_{t+i-1-m}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top \right) \right. \\
& \left. \cdot \text{vec} \left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{m,t+i-1} \right) \right) \boldsymbol{\eta}_{t+i}^\top K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right) := J_{T,32241} + J_{T,32242}.
\end{aligned}$$

Similar to the proof of $J_{T,3221}$, we can show that $J_{T,32242} = O_P((Th)^{-1/2})$.

Let $\boldsymbol{w}_{t,i} = \sum_{j=0}^{\infty} \left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top \right) \text{vec} \left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{i-1,t+i-1} \right)$. For $J_{T,32241}$,

$$\begin{aligned}
E \|J_{T,32241}\|^2 & = \frac{1}{(Th)^3} \sum_{i_1=1}^{T-1} \sum_{i_2=1}^{T-1} \sum_{t_1=1}^{T-i_1} E \left\| \boldsymbol{\eta}_{t_1+i_1}^\top \boldsymbol{\eta}_{t_1+i_1} \right\| E \left\| \boldsymbol{w}_{t_1+i_1-i_2, i_2}^\top \boldsymbol{w}_{t_1, i_1} \right\| \\
& \cdot K \left(\frac{i_1}{Th} \right) K \left(\frac{i_2}{Th} \right) K \left(\frac{\tau_{t_1} - \tau}{h} \right) K \left(\frac{\tau_{t_1+i_1-i_2} - \tau}{h} \right) \\
& \leq \frac{M}{(Th)^3} \sum_{t_1=1}^T \left(\max_t \sum_{i=1}^{T-1} \|\boldsymbol{\Psi}_{i,t}\| \right)^2 \left(\max_t \sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_{j,t}\| \right)^2 K \left(\frac{\tau_{t_1} - \tau}{h} \right) = O((Th)^{-2}).
\end{aligned}$$

Hence, $J_{T,32} = O_P((Th)^{-1/2})$. Similar to $J_{T,32}$, $J_{T,33} = O_P((Th)^{-1/2})$. The proof is now completed.

(3). By Lemma B.5, we have

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

Then we just need to focus on the rate associated with $\widehat{\boldsymbol{\Omega}}(\tau)$. For notational simplicity, we ignore the $\frac{1}{T} \sum_{t=1}^T K_h(\tau_t - \tau)$, because

$$\frac{1}{T} \sum_{t=1}^T K_h(\tau_t - \tau) = 1 + O((Th)^{-1})$$

uniformly over $\tau \in [h, 1-h]$.

Write

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

Consider $I_{T,1}$. Similar to the proof of part (1), we have

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

Next, consider $I_{T,2}$. By Lemma B.5, we have

$$\begin{aligned} &\sup_{\tau \in [0,1]} \frac{1}{T} \sum_{t=1}^T \|\mathbf{Z}_{t-1}\|^2 K_h(\tau_t - \tau) \\ &\leq \sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^T \text{tr} \left[\mathbf{Z}_{t-1}^\top \mathbf{Z}_{t-1} - E(\mathbf{Z}_{t-1}^\top \mathbf{Z}_{t-1}) \right] K_h(\tau_t - \tau) \right| \\ &\quad + \sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^T \text{tr} \left[E(\mathbf{Z}_{t-1}^\top \mathbf{Z}_{t-1}) \right] K_h(\tau_t - \tau) \right| \\ &= o_P(1) + O(1) = O_P(1). \end{aligned}$$

Hence, by the first result of Theorem 2.1

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

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

$$\begin{aligned}
\sup_{\tau \in [0,1]} \|I_{T,3}\| &\leq \sup_{\tau \in [0,1]} \left\| \widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| \cdot \sup_{\tau \in [0,1]} \frac{1}{T} \sum_{t=1}^T \|\mathbf{Z}_{t-1} \boldsymbol{\eta}_t\| K_h(\tau_t - \tau) \\
&\leq \sup_{\tau \in [0,1]} \left\| \widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\| \cdot \left\{ \sup_{\tau \in [0,1]} \frac{1}{T} \sum_{t=1}^T \|\mathbf{Z}_{t-1}\|^2 K_h(\tau_t - \tau) \right\}^{1/2} \\
&\quad \cdot \left\{ \sup_{\tau \in [0,1]} \frac{1}{T} \sum_{t=1}^T \|\boldsymbol{\eta}_t\|^2 K_h(\tau_t - \tau) \right\}^{1/2} \\
&= O_P \left(h^2 + \sqrt{\frac{\log T}{Th}} \right).
\end{aligned}$$

The proof is now completed. \square

Define $\boldsymbol{\Lambda}_p(\tau) = [\mathbf{a}(\tau), \mathbf{A}_{p,1}(\tau), \dots, \mathbf{A}_{p,p}(\tau)]$, where $\mathbf{A}_{p,j}(\tau) = \mathbf{A}_j(\tau)$ for $1 \leq j \leq p$ and $\mathbf{A}_{p,j}(\tau) = 0$ for $j > p$. Let $\mathbf{z}_{p,t-1} = [1, \mathbf{x}_{t-1}^\top, \dots, \mathbf{x}_{t-p}^\top]^\top$, $\mathbf{z}_{p,t-1}^* = [\mathbf{z}_{p,t-1}^\top, \frac{\tau_t - \tau}{h} \mathbf{z}_{p,t-1}^\top]^\top$, $\mathbf{Z}_{p,t}^* = \mathbf{z}_{p,t}^* \otimes \mathbf{I}_d$, $\mathbf{M}_p(\tau_t) = \boldsymbol{\Lambda}_p(\tau_t) - \boldsymbol{\Lambda}_p(\tau) - \boldsymbol{\Lambda}_p^{(1)}(\tau)(\tau_t - \tau) - \frac{1}{2} h^2 \boldsymbol{\Lambda}_p^{(2)}(\tau)(\tau_t - \tau)^2$, $\boldsymbol{\Lambda}_{\bar{p}}(\tau) = [\mathbf{A}_{p,p+1}(\tau), \dots, \mathbf{A}_{p,p}(\tau)]$ and $\mathbf{z}_{\bar{p},t-1} = [\mathbf{x}_{t-p-1}^\top, \dots, \mathbf{x}_{t-p}^\top]^\top$.

Proof of Lemma B.7.

(1). Since $\mathbf{p} \geq p$, we have $\widehat{\boldsymbol{\eta}}_{p,t} = \boldsymbol{\eta}_t + (\boldsymbol{\Lambda}_p(\tau_t) - \widehat{\boldsymbol{\Lambda}}_p(\tau_t)) \mathbf{z}_{p,t-1}$ and

$$\begin{aligned}
\text{RSS}(\mathbf{p}) &= \frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t + \frac{1}{T} \sum_{t=1}^T \mathbf{z}_{p,t-1}^\top (\boldsymbol{\Lambda}_p(\tau_t) - \widehat{\boldsymbol{\Lambda}}_p(\tau_t))^\top (\boldsymbol{\Lambda}_p(\tau_t) - \widehat{\boldsymbol{\Lambda}}_p(\tau_t)) \mathbf{z}_{p,t-1} \\
&\quad - 2 \frac{1}{T} \sum_{t=1}^T \text{tr} \left(\boldsymbol{\eta}_t (\boldsymbol{\eta}_t - \widehat{\boldsymbol{\eta}}_{p,t})^\top \right) := \frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t + I_{T,1} + I_{T,2}.
\end{aligned}$$

Since $\boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t - E(\boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t)$ is m.d.s., we have $\frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t = \frac{1}{T} \sum_{t=1}^T E(\boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t) + T^{-1/2}$. By Theorem 2.1.1,

$$\begin{aligned}
I_{T,1} &\leq \frac{1}{T} \sum_{t=1}^T \|\mathbf{z}_{p,t-1}\|^2 \cdot \left\| \widehat{\boldsymbol{\Lambda}}_p(\tau_t) - \boldsymbol{\Lambda}_p(\tau_t) \right\|^2 \leq \sup_{0 \leq \tau \leq 1} \left\| \widehat{\boldsymbol{\Lambda}}_p(\tau) - \boldsymbol{\Lambda}_p(\tau) \right\|^2 \cdot \frac{1}{T} \sum_{t=1}^T \|\mathbf{z}_{p,t-1}\|^2 \\
&= O_P \left((h^2 + (\log T / (Th))^{1/2})^2 \right).
\end{aligned}$$

For $I_{T,2}$,

$$\begin{aligned}
&\frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t (\boldsymbol{\eta}_t - \widehat{\boldsymbol{\eta}}_{p,t})^\top = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t \mathbf{z}_{p,t-1}^\top (\widehat{\boldsymbol{\Lambda}}_p(\tau_t) - \boldsymbol{\Lambda}_p(\tau_t))^\top \\
&= \frac{1}{2} h^2 \cdot \frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t [\mathbf{z}_{p,t-1}^\top, \mathbf{0}_{(d^2 p + d) \times 1}^\top] \mathbf{S}_T^{-1}(\tau_t) \begin{pmatrix} \mathbf{S}_{T,2}(\tau_t) \\ \mathbf{S}_{T,3}(\tau_t) \end{pmatrix} \mathbf{A}_p^{(2),\top}(\tau_t) \\
&+ \frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t [\mathbf{z}_{p,t-1}^\top, \mathbf{0}_{(d^2 p + d) \times 1}^\top] \mathbf{S}_T^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^T (\mathbf{z}_{p,s-1}^* \mathbf{z}_{p,s-1}^\top \otimes \mathbf{I}_d) \mathbf{M}_p^\top(\tau_s) K \left(\frac{\tau_s - \tau_t}{h} \right) \right) \\
&+ \frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t [\mathbf{z}_{p,t-1}^\top, \mathbf{0}_{(d^2 p + d) \times 1}^\top] \mathbf{S}_T^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^T \mathbf{Z}_{p,s-1}^* \boldsymbol{\eta}_s^\top K \left(\frac{\tau_s - \tau_t}{h} \right) \right) := I_{T,3} + I_{T,4} + I_{T,5}.
\end{aligned}$$

By the uniform convergence results stated in Lemmas B.5.2-3, we replace the weighed sample covariance with its limit plus the rate $O_P((\log T/(Th))^{1/2})$, and hence

$$\|I_{T,3}\| + \|I_{T,4}\| = O_P\left(T^{-\frac{1}{2}}h^2 + h^2(\log T/(Th))^{1/2}\right).$$

For $I_{T,5}$, let $\boldsymbol{\Sigma}(\tau) = \text{plim}_{T \rightarrow \infty} \mathbf{S}_{T,0}(\tau)$, we have

$$I_{T,6} = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t \mathbf{z}_{\mathbf{p},t-1}^\top \boldsymbol{\Sigma}^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^T \mathbf{z}_{\mathbf{p},s-1} \boldsymbol{\eta}_s^\top K\left(\frac{\tau_s - \tau_t}{h}\right) \right) + O_P\left((Th)^{-1/2} \cdot (h^2 + \sqrt{\log T/(Th)})\right).$$

Similar to the proof of $J_{T,4}$ in Lemma B.6, we can show

$$\frac{1}{T} \sum_{t=1}^T \boldsymbol{\eta}_t \mathbf{z}_{\mathbf{p},t-1}^\top \boldsymbol{\Sigma}^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^T \mathbf{z}_{\mathbf{p},s-1} \boldsymbol{\eta}_s^\top K\left(\frac{\tau_s - \tau_t}{h}\right) \right) = O_P((Th)^{-1}).$$

Since $(Th)^{-1} + T^{-\frac{1}{2}}h^{\frac{3}{2}} = o(\rho_T^2)$, result (1) follows.

(2). For $\mathbf{p} < p$, we have $\widehat{\boldsymbol{\Lambda}}_{\mathbf{p}}(\tau) - \boldsymbol{\Lambda}_{\mathbf{p}}(\tau) = \mathbf{B}_{\mathbf{p}}(\tau) + o_P(1)$ uniformly over $\tau \in [0, 1]$, where $\mathbf{B}_{\mathbf{p}}(\tau)$ is a nonrandom bias term. Since $\widehat{\boldsymbol{\eta}}_{\mathbf{p},t} = \boldsymbol{\eta}_t + (\boldsymbol{\Lambda}_{\mathbf{p}}(\tau_t) - \widehat{\boldsymbol{\Lambda}}_{\mathbf{p}}(\tau_t)) \mathbf{z}_{\mathbf{p},t-1} + \boldsymbol{\Lambda}_{\bar{\mathbf{p}}}(\tau_t) \mathbf{z}_{\bar{\mathbf{p}},t-1}$, by Lemma B.5.4, we have

$$\text{RSS}(\mathbf{p}) = \frac{1}{T} \sum_{t=1}^T E\left(\boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t\right) + \frac{1}{T} \sum_{t=1}^T \text{tr}\left([\mathbf{B}_{\mathbf{p}}(\tau_t), \boldsymbol{\Lambda}_{\bar{\mathbf{p}}}(\tau_t)] E\left(\mathbf{z}_{\mathbf{p},t-1} \mathbf{z}_{\bar{\mathbf{p}},t-1}^\top\right) [\mathbf{B}_{\mathbf{p}}(\tau_t), \boldsymbol{\Lambda}_{\bar{\mathbf{p}}}(\tau_t)]^\top\right) + o_P(1).$$

Since $[\mathbf{B}_{\mathbf{p}}(\tau_t), \boldsymbol{\Lambda}_{\bar{\mathbf{p}}}(\tau_t)] \neq 0$ and $E\left(\mathbf{z}_{\mathbf{p},t-1} \mathbf{z}_{\bar{\mathbf{p}},t-1}^\top\right)$ is a positive definite matrix, the result follows. \square

Proof of Lemma B.8.

(1). Write

$$\begin{aligned} \sigma^2 &= \frac{1}{T^2 h} \sum_{t=2}^T \sum_{s=1}^{t-1} \left[\int_{-1}^1 K(u) K\left(u + \frac{t-s}{Th}\right) du \right]^2 \\ &= \frac{1}{T^2 h} \sum_{t=2}^T \sum_{j=1}^{t-1} \left[\int_{-1}^1 K(u) K\left(u + \frac{j}{Th}\right) du \right]^2 \\ &= \frac{1}{Th} \sum_{j=1}^{T-1} (1 - j/T) \left[\int_{-1}^1 K(u) K\left(u + \frac{j}{Th}\right) du \right]^2 \\ &= \int_0^\infty (1 - vh) \left[\int_{-1}^1 K(u) K(u+v) du \right]^2 dv + O(1/(Th)) \\ &\rightarrow \int_0^2 \left[\int_{-1}^{1-v} K(u) K(u+v) du \right]^2 dv. \end{aligned}$$

(2). Write

$$\begin{aligned} \max_t |a_t| &= \max_t \left| \sum_{i=1}^{t-1} \frac{1}{T^2 h} \left[\int_{-1}^1 K(u) K\left(u + \frac{i}{Th}\right) du \right]^2 \right| \\ &\leq \frac{1}{T^2 h} \sum_{i=1}^T \left[\int_{-1}^1 K(u) K\left(u + \frac{i}{Th}\right) du \right]^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{T} \int_0^\infty \left[\int_{-1}^1 K(u) K(u+v) du \right]^2 dv (1 + o(1)) \\
&= O(1/T).
\end{aligned}$$

(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

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

(5). By Cauchy-Schwarz inequality,

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

□

Proof of Lemma B.9.

Write $\xi_t = [\xi_{t,1}, \dots, \xi_{t,d}]^\top$. We first prove

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

for $1 \leq i \leq d$.

By Burkholder inequality, the Minkowski inequality and the inequality that $|\sum_{i=1}^n a_i|^p \leq \sum_{i=1}^n |a_i|^p$ for $0 < p \leq 1$, we have

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

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

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

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

Proof of Lemma B.10.

Without loss of generality, let $E(\mathbf{w}_t^*) = \mathbf{0}$. For any integer $I \geq 1$ introduce the truncated process $\mathbf{h}_{t-1,I}^* = E(\mathbf{h}_{t-1}^* | \mathcal{F}_{t-I})$. Then $\mathbf{h}_{t-1,I}^* = \mathbf{0}$ if $t \leq I$ and $\mathbf{h}_{t-1,I}^* = \sum_{s=1}^{t-I} w_{s,t} \mathbf{y}_s$ for $1 \leq I < t$. For $2 \leq t \leq T$, by Lemma B.9,

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

Let $L(I) = \sum_{J=1}^I l(J)$ with $l(J) = \sum_{s=1}^{T-J} w_{s,s+J}^2$, $V(I) = \sum_{t=2}^T \text{tr} \left[\mathbf{w}_t^* \mathbf{h}_{t-1,I}^* \mathbf{h}_{t-1,I}^{*\top} \right]$ and $T(I) = \sum_{t=2}^T \text{tr} \left[E(\mathbf{w}_t^* | \mathcal{F}_{t-I}) \right]$

By Cauchy-Schwarz inequality, Lemma B.8 (iii), if $I/(Th) \rightarrow 0$, we have

$$\begin{aligned} E|V(1) - V(I)| &\leq \sum_{t=2}^T E \left| \text{tr} \left[\mathbf{w}_t^* \left(\mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top} - \mathbf{h}_{t-1,I}^* \mathbf{h}_{t-1,I}^{*\top} \right) \right] \right| \\ &\leq \sum_{t=2}^T \|\mathbf{w}_t^*\| \|\mathbf{h}_{t-1}^* - \mathbf{h}_{t-1,I}^*\|_4 \|\mathbf{h}_{t-1}^* + \mathbf{h}_{t-1,I}^*\|_4 \\ &\leq M \sum_{t=2}^T \|\mathbf{h}_{t-1}^* - \mathbf{h}_{t-1,I}^*\|_4 a_t^{1/2} \\ &\leq M \left\{ \sum_{t=2}^T \|\mathbf{h}_{t-1}^* - \mathbf{h}_{t-1,I}^*\|_4^2 \right\}^{1/2} \left\{ \sum_{t=2}^T a_t \right\}^{1/2} \\ &= O(1)[L(I)]^{1/2} \rightarrow 0, \end{aligned}$$

since

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

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

$$V(I) - T(I) = \sum_{t=2}^T \text{tr} \left[\left(\sum_{j=0}^{I-1} \mathcal{P}_{t-j} \mathbf{w}_t^* \right) \mathbf{h}_{t-1,I}^* \mathbf{h}_{t-1,I}^{*\top} \right] = \sum_{j=0}^{I-1} U(j).$$

Note that $\left\{ (\mathcal{P}_{t-j} \mathbf{w}_t^*) \mathbf{h}_{t-1,I}^* \mathbf{h}_{t-1,I}^{*\top} \right\}_{t=2}^T$ forms a martingale difference sequence since

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

By Lemma B.8 (ii), Lemma B.9 and Cauchy-Schwarz inequality, since $\left\| \mathcal{P}_{t-j} \mathbf{w}_t^* \right\|_{\delta/2} \leq 2 \left\| \mathbf{w}_t^* \right\|_{\delta/2} < \infty$,

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

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

$$\begin{aligned} E|V(1)| &\leq \left\| V(I) - T(I) \right\|_{\delta/4} + E|T(I)| + o(1) \\ &\leq \sum_{j=0}^{I-1} \left\| U(j) \right\|_{\delta/4} + \max_t \left\| E(\mathbf{w}_t^* | \mathcal{F}_{t-I}) \right\| \sum_{t=2}^T \left\| \mathbf{h}_{t-1,I}^* \right\|_4^2 = o(1), \end{aligned}$$

since $\max_t \left\| E(\mathbf{w}_t^* | \mathcal{F}_{t-I}) \right\| \rightarrow 0$ as $I \rightarrow \infty$. The proof is now completed. \square

Proof of Lemma B.11.

For notational simplicity, let $\mathbf{H}_t = \mathbf{I}_{d^2}$. Write

$$\begin{aligned} &\sum_{t=2}^T \text{tr} \left[\mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top} - E \left(\mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top} \right) \right] \\ &= \sum_{t=2}^T \sum_{s=1}^{t-1} \text{tr} \left[\left(\mathbf{Z}_{s-1} \boldsymbol{\eta}_s \boldsymbol{\eta}_s^\top \mathbf{Z}_{s-1}^\top - E \left(\mathbf{Z}_{s-1} \boldsymbol{\eta}_s \boldsymbol{\eta}_s^\top \mathbf{Z}_{s-1}^\top \right) \right) w_{s,t}^2 \right] \\ &\quad + 2 \sum_{t=3}^T \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} \text{tr} \left[\mathbf{Z}_{s_1-1} \boldsymbol{\eta}_{s_1} \boldsymbol{\eta}_{s_2}^\top \mathbf{Z}_{s_2-1}^\top w_{s_1,t} w_{s_2,t} \right] \\ &= I_{T,1} + 2I_{T,2}. \end{aligned}$$

Consider $I_{T,1}$. Write

$$\begin{aligned}
I_{T,1} &= \sum_{t=2}^T \sum_{s=1}^{t-1} \text{tr} \left[\left(\boldsymbol{\eta}_s \boldsymbol{\eta}_s^\top - \boldsymbol{\Omega}(\tau_s) \right) \mathbf{Z}_{s-1}^\top \mathbf{Z}_{s-1} \right] w_{s,t}^2 \\
&\quad + \sum_{t=2}^T \sum_{s=1}^{t-1} \text{tr} \left[\boldsymbol{\Omega}(\tau_s) \left(\mathbf{Z}_{s-1}^\top \mathbf{Z}_{s-1} - E \left(\mathbf{Z}_{s-1}^\top \mathbf{Z}_{s-1} \right) \right) \right] w_{s,t}^2 \\
&= \frac{1}{T} \sum_{s=1}^{T-1} \text{tr} \left[\left(\boldsymbol{\eta}_s \boldsymbol{\eta}_s^\top - \boldsymbol{\Omega}(\tau_s) \right) \mathbf{Z}_{s-1}^\top \mathbf{Z}_{s-1} \right] \left(T \sum_{t=s+1}^T w_{s,t}^2 \right) \\
&\quad + \frac{1}{T} \sum_{s=1}^{T-1} \text{tr} \left[\boldsymbol{\Omega}(\tau_s) \left(\mathbf{Z}_{s-1}^\top \mathbf{Z}_{s-1} - E \left(\mathbf{Z}_{s-1}^\top \mathbf{Z}_{s-1} \right) \right) \right] \left(T \sum_{t=s+1}^T w_{s,t}^2 \right) \\
&= I_{T,11} + I_{T,12}.
\end{aligned}$$

Since $\text{tr} \left[\left(\boldsymbol{\eta}_s \boldsymbol{\eta}_s^\top - \boldsymbol{\Omega}(\tau_s) \right) \mathbf{Z}_{s-1}^\top \mathbf{Z}_{s-1} \right]$, $s = 1, 2, \dots$ are a martingale difference sequence and $T \sum_{t=1}^T w_{s,t}^2 = O(1)$ by Lemma B.8.2, we have $I_{T,11} = o_P(1)$. In addition, by Lemma B.5.4, we have $I_{T,12} = o_P(1)$.

Next, consider $I_{T,2}$. By Lemma B.9, Cauchy-Schwarz inequality and Lemma B.8.5,

$$\begin{aligned}
\|I_{T,2}\|^2 &\leq M \sum_{s_1=2}^{T-1} \left\| \text{tr} \left[\mathbf{Z}_{s_1-1} \boldsymbol{\eta}_{s_1} \sum_{s_2=1}^{s_1-1} \boldsymbol{\eta}_{s_2} \mathbf{Z}_{s_2-1}^\top \sum_{t=s_1+1}^T w_{s_1,t} w_{s_2,t} \right] \right\|^2 \\
&\leq M \sum_{s_1=2}^{T-1} \|\mathbf{Z}_{s_1-1} \boldsymbol{\eta}_{s_1}\|_4^2 \left\| \sum_{s_2=1}^{s_1-1} \mathbf{Z}_{s_2-1} \boldsymbol{\eta}_{s_2} \sum_{t=s_1+1}^T w_{s_1,t} w_{s_2,t} \right\|_4^2 \\
&\leq M \sum_{s_1=2}^{T-1} \|\mathbf{Z}_{s_1-1} \boldsymbol{\eta}_{s_1}\|_4^2 \sum_{s_2=1}^{s_1-1} \left\| \mathbf{Z}_{s_2-1} \boldsymbol{\eta}_{s_2} \sum_{t=s_1+1}^T w_{s_1,t} w_{s_2,t} \right\|_4^2 \\
&= O \left(\sum_{s_1=2}^{T-1} \sum_{s_2=1}^{s_1-1} \left(\sum_{t=s_1+1}^T w_{s_1,t} w_{s_2,t} \right)^2 \right) = O(1/T).
\end{aligned}$$

Combine the above results, the proof is now complete. \square

Proof of Lemma B.12. Note that $\mathbf{y}_t^\top \mathbf{H}_t \mathbf{h}_{t-1}^*$, $t \in \mathbb{Z}$, are martingale differences with respect to the filtration \mathcal{F}_t . We apply Lemma B.1 to prove the asymptotic normality of Q_T . By $E \left(\|\boldsymbol{\epsilon}_t\|^\delta | \mathcal{F}_{t-1} \right) < \infty$ a.s., we have

$$E \left[\|\mathbf{y}_t\|^\delta \right] \leq E \left[E \left(\|\boldsymbol{\epsilon}_t\|^\delta | \mathcal{F}_{t-1} \right) \|\mathbf{z}_{t-1} \otimes \mathbf{I}_d\|^\delta \right] < E \left[M \|\mathbf{z}_{t-1}\|^\delta \right] < \infty.$$

By Cauchy-Schwarz inequality, Lemma B.9 and Lemma B.8 (2), the Lindeberg condition is satisfied since

$$\begin{aligned}
\sum_{t=2}^T \left\| \mathbf{y}_t \mathbf{H}_t \mathbf{h}_{t-1}^* \right\|_{\delta/2}^{\delta/2} &\leq \sum_{t=2}^T \left\| \mathbf{y}_t \mathbf{H}_t \right\|_{\delta}^{\delta/2} \left\| \mathbf{h}_{t-1}^* \right\|_{\delta}^{\delta/2} \\
&\leq M \max_t \left\| \mathbf{y}_t \right\|_{\delta}^{\delta} \sum_{t=2}^T a_t^{\delta/4} \\
&= O(1) \cdot \max_t a_t^{\delta/4-1} = o(1).
\end{aligned}$$

Apply Lemmas B.10 and B.11 with $\mathbf{w}_t^* = E \left(\mathbf{H}_t^\top \mathbf{y}_t \mathbf{y}_t^\top \mathbf{H}_t | \mathcal{F}_{t-1} \right)$, then we have the convergence of condi-

tional variance

$$\begin{aligned} & \sum_{t=2}^T \text{tr} \left[E \left(\mathbf{H}_t^\top \mathbf{y}_t \mathbf{y}_t^\top \mathbf{H}_t | \mathcal{F}_{t-1} \right) \mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top} \right] \\ & \rightarrow_P \sum_{t=2}^T \text{tr} \left[E \left(\mathbf{H}_t^\top \mathbf{y}_t \mathbf{y}_t^\top \mathbf{H}_t \right) E \left(\mathbf{h}_{t-1}^* \mathbf{h}_{t-1}^{*\top} \right) \right]. \end{aligned}$$

By Lemma B.1, the proof is now completed. □

References

- Breusch, T. S. (1978), ‘Testing for autocorrelation in dynamic linear models’, *Australian Economic Papers* **17**(31), 334–355.
- Cheng, T., Gao, J. and Zhang, X. (2019), ‘A Bayesian approach in a nonparametric time-varying regression model’, *Journal of Business and Economic Statistics* **37**(1), 1–12.
- Cogley, T., Primiceri, G. E. and Sargent, T. J. (2010), ‘Inflation-gap persistence in the U.S.’, *American Economic Journal: Macroeconomics* **2**(1), 43–69.
- Cogley, T. and Sargent, T. J. (2005), ‘Drifts and volatilities: Monetary policies and outcomes in the post World War II U.S.’, *Review of Economic Dynamics* **8**(2), 262–302.
- Connor, G., Hagmann, M. and Linton, O. (2012), ‘Efficient semiparametric estimation of the Fama-French model and extensions’, *Econometrica* **80**(2), 713–754.
- Dahlhaus, R. (1996), ‘On the kullback-leibler information divergence of locally stationary processes’, *Stochastic Processes and Their Applications* **62**(1), 139–168.
- Dahlhaus, R. and Polonik, W. (2009), ‘Empirical spectral processes for locally stationary time series’, *Bernoulli* **15**(1), 1–39.
- Dahlhaus, R. and Rao, S. S. (2006), ‘Statistical inference for time-varying ARCH processes’, *Annals of Statistics* **34**(3), 1075–1114.
- Elbourne, A. and de Haan, J. (2006), ‘Financial structure and monetary policy transmission in transition countries’, *Journal of Comparative Economics* **34**(1), 1–23.
- 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.
- Giraitis, L., Kapetanios, G. and Yates, T. (2014), ‘Inference on stochastic time-varying coefficient models’, *Journal of Econometrics* **179**(1), 46–65.
- Godfrey, L. G. (1978), ‘Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables’, *Econometrica* **46**(6), 1293–1301.
- Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Academic Press.

- Hansen, B. E. (2001), ‘The new econometrics of structural change: dating breaks in US labour productivity’, *Journal of Economic Perspectives* **15**(4), 117–128.
- Hong, Y. and Li, H. (2005), ‘Nonparametric specification testing for continuous-time models with applications to term structure of interest rates’, *Review of Financial Studies* **18**(1), 37–84.
- Inoue, A. and Kilian, L. (2013), ‘Inference on impulse response functions in structural var models’, *Journal of Econometrics* **117**(1), 1–13.
- Inoue, A. and Kilian, L. (2020), ‘The uniform validity of impulse response inference in autoregressions’, *Journal of Econometrics* **215**(3), 450–472.
- Jentsch, C. and Lunsford, K. G. (2021), ‘Asymptotically valid bootstrap inference for proxy svars’, *Journal of Business & Economic Statistics* p. forthcoming.
- Kilian, L. and Lütkepohl, H. (2017), *Structural Vector Autoregressive Analysis*, Cambridge University Press.
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.
- Lütkepohl, H. and Schlaak, T. (2021), ‘Heteroscedastic proxy vector autoregressions’, *Journal of Business & Economic Statistics* p. forthcoming.
- Paul, P. (2020), ‘The time-varying effect of monetary policy on asset prices’, *Review of Economics and Statistics* **102**(4), 690–704.
- Petrova, K. (2019), ‘A quasi-Bayesian local likelihood approach to time varying parameter VAR models’, *Journal of Econometrics* **212**(1), 286–306.
- Primiceri, G. E. (2005), ‘Time varying structural vector autoregressions and monetary policy’, *Review of Economic Studies* **72**(3), 821–852.
- Richter, S. and Dahlhaus, R. (2019), ‘Cross validation for locally stationary processes’, *Annals of Statistics* **47**(4), 2145–2173.
- Rusnák, M., Havranek, T. and Horváth, R. (2013), ‘How to solve the price puzzle? a meta-analysis’, *Journal of Money, Credit and Banking* **45**(1), 37–70.
- Sims, C. A. (1980), ‘Macroeconomics and reality’, *Econometrica* **48**(1), 1–48.
- Sims, C. and Zha, T. (2006), ‘Were there regime switches in U.S. monetary policy?’, *American Economic Review* **96**(1), 54–81.
- Stock, J. H. and Watson, M. W. (2001), ‘Vector autoregressions’, *Journal of Economic Perspectives* **15**(4), 101–115.
- Stock, J. H. and Watson, M. W. (2016), Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics, in ‘Handbook of Macroeconomics’, Vol. 2, Elsevier, pp. 415–525.
- Sun, Y., Hong, Y., Lee, T. H., Wang, S. and Zhang, X. (2021), ‘Time-varying model averaging’, *Journal of Econometrics* **222**(2), 974–992.
- 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 modelling multivariate threshold models’, *Journal of the American Statistical Association* **93**(443), 1188–1202.
- Vogt, M. (2012), ‘Nonparametric regression for locally stationary time series’, *Annals of Statistics* **40**(5), 2601–2633.
- 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.