MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Nonparametric Estimation and Testing for Time- Varying VAR models

Jiti Gao
Bin Peng
Yayi Yan

Working Paper No. 03/22
March 2022

Nonparametric Estimation and Testing for Time–Varying VAR Models

Jiti Gao, Bin Peng and Yayi Yan

Monash University

March 21, 2022

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 asymptotic properties including the impulse response analyses subject to structural VAR identification conditions, 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 on US government spending multipliers.

Keywords: Time–Varying Impulse Response; Parameter Stability; Instrumental Variable Approach

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 and Peng acknowledge financial support from the Australian Research Council Discovery Grants Program under Grant Numbers: DP170104421 and DP210100476, respectively.

1 Introduction

Vector autoregressive (VAR) models as well as their extensions are among some of the most popular frameworks for modelling dynamic interactions of 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 of policy analysis (Stock and Watson, 2001). Despite the popularity, linear VAR models can always be rejected by data in empirical applications (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 better model the dynamic transit, one strand of the VAR literature considers stochastic time–varying coefficients involved in VAR models (e.g., Primiceri, 2005). The estimation methods proposed in this strand of literature basically rely on extensive Markov Chain Monte Carlo (MCMC) draws plus the use of a variety of filters, such as Kalman and related filters. As pointed out by Giraitis et al. (2014), however, this strand of the VAR literature hasn't paid attention on asymptotic justifications and properties for the estimated model coefficients as well as the corresponding impulse responses.

Another strand of relevant literature focuses on nonparametric estimation for deterministic time–varying coefficients involved in autoregressive models. Up to this point, it is worth bringing up the terminology "local stationarity", which dates back at least to the seminal work of Dahlhaus (1996). Locally stationary processes are useful when analysing economic and financial time series (Sun et al., 2021; Xu et al., 2021). For example, the dataset plotted in Figure 1 of Section 5 shows that each univariate time series is globally nonstationary, but locally stationary. There have since been developments, such as Dahlhaus and Rao (2006), Zhang and Wu (2012), Richter and Dahlhaus (2019) and Yan et al. (2021). Meanwhile, there is a related strand of literature on nonparametric estimation for time–varying parameters for time series regression models (Robinson, 1989; Gao and Hawthorne, 2006; Cai, 2007; Chen and Hong, 2012; Phillips et al., 2017; Li et al., 2020). It should be pointed out that this strand of literature is not relevant to what we discuss in this paper where time–varying parameters are involved in dynamic systems, including VAR models.

In view of the aforementioned issues, this paper therefore investigates a class of time–varying VAR models where both VAR coefficients and covariance matrix of the model's error innovations are allowed to evolve over time. Such modelling strategy is especially useful for analysing multivariate time series over a long horizon, as it helps track frequently updated policies, environment, system, etc. (Hansen, 2001; Phillips et al., 2017). In fact, a few attempts sharing the similar motivations have been made in recent years. For example, Giraitis et al. (2018) and Kapetanios et al. (2019) consider a stochastic time–varying framework using a set of high level conditions, but no results on impulse responses are provided. And yet, it is hard to justify the high level conditions in practical applications. Moreover, there is no statistical evidence to support whether their approaches should be preferred to a typical parametric model.

That said, in this paper, we specifically study the asymptotic properties of the impulse responses under different identification conditions, which are widely adopted in the literature for different purposes. Also, we provide the statistical support to help researchers decide when a time–varying framework should be preferred in practice. In order to achieve these, from the methodological viewpoint, we first develop a time–varying vector moving average infinity (VMA(∞)) representation for a class of VAR models. Then we establish uniform consistency and a joint central limit theory for the estimators of VAR coefficients and covariance matrix. Afterwards, we derive the asymptotic properties of the time–varying impulse responses, which are of importance in typical VAR applications (Inoue and Kilian, 2013, 2020; Paul, 2020). A few identifications conditions (e.g., structural VAR (SVAR) identification schemes and external IV method) are considered in order to broaden the applicability of the newly proposed framework. Last but not least, we establish a hypothesis test to examine the parameter stability, which provides statistical evidence to support the necessity of the time–varying VAR models for real data applications.

Up to this point, we briefly comment on the literature of parameter stability test, which has also received considerable attention over the past ten years. Such a test for example can be used to examine whether the policy transmission mechanism is changing with respect to time (Primiceri, 2005; Paul, 2020). Detecting parametric components in univariate time–varying autoregressive models is studied in Zhang and Wu (2012), which is further extended by Truquet (2017) to examine time–varying autoregressive conditional heteroscedasticity. In this paper, we specifically develop an integrated L_2 type test for checking whether some (if not all) of the coefficients are constant.

Finally, in the empirical study, we use the newly established framework and results to investigate the US government spending multipliers. We find that the government spending multipliers are above one before 1990s and are not significantly from zero after 1990s, which is consistent using different identification schemes. In a sense, our finding provides numerical support to Ramey and Zubairy (2018), who argue that "*Increases over time in financial market access and consumer sophistication should reduce the fraction of rule-of-thumb consumers, thus reducing multipliers in recent years.*"

The organization of this paper is as follows. Section 2 considers a class of time–varying VAR models, and establishes the corresponding asymptotics. Section 3 specifically focuses on inferring the time–varying impulse responses subject to different identification schemes. Section 4 discusses some implementation issues and presents comprehensive simulations. Section 5 provides a case study to demonstrate the empirical relevance. Section 6 concludes. Some lengthy mathematical symbols are summarized in Appendix A. The preliminary lemmas and the proofs are given in Appendix B of the online supplementary document.

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; $\boldsymbol{I}_a$ 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 $\tilde{c}_k=\int_{-1}^{1}u^kK(u)du$ and $\tilde{v}_k=\int_{-1}^{1}u^kK^2(u)du$ 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 $\boldsymbol{A}$, $\text{vech}\,(\boldsymbol{A})$ denotes the $\frac{1}{2}a(a+1)\times 1$ vector obtained from $\text{vec}\,(\boldsymbol{A})$ by eliminating all supra–diagonal elements of $\boldsymbol{A}$; $\text{tr}\,(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$; let $\boldsymbol{A}_{i,\boldsymbol{\cdot}}$ and $\boldsymbol{A}_{\boldsymbol{\cdot},j}$ denote the i^{th} row and the j^{th} column of matrix $\boldsymbol{A}$ respectively; finally, $\to_P$ and $\to_D$ denote convergence in probability and convergence in distribution.

2 The Time–Varying VAR(p) Model

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

$$\boldsymbol{x}_t=\boldsymbol{a}(\tau_t)+\sum_{j=1}^{p}\boldsymbol{A}_j(\tau_t)\boldsymbol{x}_{t-j}+\boldsymbol{\eta}_t, \tag{2.1}$$

where $\tau_t=t/T$, $\boldsymbol{\eta}_t=\boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t$, and $\boldsymbol{a}(\cdot)$, $\boldsymbol{A}_j(\cdot)$ and $\boldsymbol{\omega}(\cdot)$ are respectively $d\times 1$, $d\times d$ and $d\times d$. 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 $\boldsymbol{x}_t$ has a time–invariant effect on the j^{th} variable of $\boldsymbol{x}_t$, restricting simultaneous interactions among multiple variables to be time–invariant. For the time being, we assume p is known, and shall come back to its estimation in the end of Section 2.1.

The following conditions are necessary for our development.

Assumption 1.

1. $\boldsymbol{I}_d-\boldsymbol{A}_1(\tau)L-\cdots-\boldsymbol{A}_p(\tau)L^p\neq\mathbf{0}_d$ *for all* $\tau\in[0,1]$ *and all* $0<|L|\leq 1+\nu$ *for some*

$\nu > 0$. *Each element of* $\boldsymbol{A}(\tau) = [\boldsymbol{a}(\tau), \boldsymbol{A}_1(\tau), \ldots, \boldsymbol{A}_p(\tau)]$ *is second order continuously differentiable on* $[0,1]$ *and* $\boldsymbol{A}(\tau) = \boldsymbol{A}(0)$ *for* $\tau < 0$.

2. *Each element of* $\boldsymbol{\omega}(\tau)$ *is second order continuously differentiable on* $[0,1]$. *Moreover,* $\boldsymbol{\Omega}(\tau) = \boldsymbol{\omega}(\tau)\boldsymbol{\omega}(\tau)^\top > 0$ *is 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* $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots)$ *is the* σ*-field generated by* $(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots)$, $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top | \mathcal{F}_{t-1}] = \boldsymbol{I}_d$ *almost surely (a.s.), and* $\max_t E\,\|\boldsymbol{\epsilon}_t\|^\delta < \infty$ *for some* $\delta > 4$.

Assumption 1.1 infers the local stationarity of (2.1). Similar treatments have also been adopted for univariate locally stationary models in the literature (e.g., Assumption T3 of Zhang and Wu, 2012). Assumption 1 also allows the underlying data generating process to evolve over time in a smooth manner. The conditions $\boldsymbol{A}(\tau) = \boldsymbol{A}(0)$ and $\boldsymbol{\omega}(\tau) = \boldsymbol{\omega}(0)$ for $\tau < 0$ gives

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

which basically assumes that $\boldsymbol{x}_t$ behaves like a parametric VAR(p) model for $t \leq 0$. A similar condition can be found in Vogt (2012) for a nonparametric time series model. Assumption 2 imposes some conditions on the innovation error terms, which are standard in the VAR literature (Lütkepohl, 2005).

Under these conditions, (2.1) admits a time–varying VMA(∞) representation, which sheds light on how to recover the time–varying structural impulse responses. Formally, we present the following proposition.

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

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

such that $\max_{t\geq 1}\{E\,\|\boldsymbol{x}_t - \widetilde{\boldsymbol{x}}_t\|^\delta\}^{1/\delta} = O(T^{-1})$, *where* $\boldsymbol{\mu}(\tau) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j(\tau)\boldsymbol{a}(\tau)$, $\boldsymbol{\Psi}_j(\tau) = \boldsymbol{J}\boldsymbol{\Phi}^j(\tau)\boldsymbol{J}^\top$, $\boldsymbol{J} = \left[\boldsymbol{I}_d, \boldsymbol{0}_{d\times d(p-1)}\right]$, $\boldsymbol{B}_j(\tau) = \boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}(\tau)$, *and* $\boldsymbol{\Phi}(\tau)$ *is defined in* (A.6) *for the sake of presentation.*

From Proposition 2.1, it is clear 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 $\boldsymbol{x}_{t+n}$ is given by $\boldsymbol{B}_n(\tau_{t+n})\boldsymbol{e}_j$, where $\boldsymbol{e}_j$ 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 $\{\boldsymbol{B}_j(\tau) : j \geq 0\}$ via the SVAR identification schemes, 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 $\boldsymbol{A}(\cdot)$ and $\boldsymbol{\Omega}(\cdot)$ using the local linear kernel method. Intuitively, when τ_t is in a small neighbourhood of τ, we can write (2.1) as follows:

$$\boldsymbol{x}_t \approx \left[\boldsymbol{A}(\tau), h\boldsymbol{A}^{(1)}(\tau)\right] \boldsymbol{z}_{t-1}^* + \boldsymbol{\eta}_t, \tag{2.3}$$

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

$$\begin{aligned}
\text{vec}[\widehat{\boldsymbol{A}}(\tau)] &= [\boldsymbol{I}_{d^2p+d}, \boldsymbol{0}_{d^2p+d}] \cdot \left(\sum_{t=1}^T \boldsymbol{Z}_{t-1}^* \boldsymbol{Z}_{t-1}^{*,\top} K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^T \boldsymbol{Z}_{t-1}^* \boldsymbol{x}_t K_h(\tau_t - \tau), \\
\widehat{\boldsymbol{\Omega}}(\tau) &= \frac{1}{T}\sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top \omega_t(\tau),
\end{aligned} \tag{2.4}$$

where $\boldsymbol{Z}_t^* = \boldsymbol{z}_t^* \otimes \boldsymbol{I}_d$, $\widehat{\boldsymbol{\eta}}_t = \boldsymbol{x}_t - \widehat{\boldsymbol{A}}(\tau_t)\boldsymbol{z}_{t-1}$, $\omega_t(\tau) = K_h(\tau_t - \tau)\frac{P_{h,2}(\tau) - \frac{\tau_t-\tau}{h}P_{h,1}(\tau)}{P_{h,0}(\tau)P_{h,2}(\tau) - P_{h,1}^2(\tau)}$ 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 $k = 0, 1, 2$.

To facilitate the development, 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 $\int_{-1}^1 K(u)\mathrm{d}u = 1$. Moreover, $K(\cdot)$ is Lipschitz continuous on $[-1, 1]$. As $(T, h) \to (\infty, 0)$, $Th \to \infty$.*

Assumption 3 is a set of regular conditions on the kernel function and the bandwidth. With these conditions 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 $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$. Then*

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

[1] It is worth pointing out that the estimation of the covariance matrix using the local linear kernel method (such as the second estimator of (2.4)) is a non-trivial problem, and even has its own literature. We refer interested readers to Zhang and Wu (2012) for more details.

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 $Th^5 \to \alpha \in [0, \infty)$. Then for $\forall \tau \in (0,1)$:

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

where $\boldsymbol{V}(\tau)$ and $\widehat{\boldsymbol{V}}(\tau)$ 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 $\widehat{\boldsymbol{A}}(\tau)$, which further allows us to establish a joint asymptotic distribution in the second result. If $\delta > 5$, the usual optimal bandwidth $h_{opt} = O\left(T^{-1/5}\right)$ satisfies the condition $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$.

We now consider the choice of the number of lags (i.e., the estimation of p). Specifically, we consider the minimization as follows:

$$\widehat{\mathsf{p}} = \underset{1\le \mathsf{p}\le \mathsf{P}}{\text{argmin}} \left(\log\{\text{RSS}(\mathsf{p})\} + \mathsf{p}\cdot\chi_T\right) \tag{2.5}$$

where $\text{RSS}(\mathsf{p}) = \frac{1}{T}\sum_{t=1}^T \widehat{\boldsymbol{\eta}}_{\mathsf{p},t}^\top \widehat{\boldsymbol{\eta}}_{\mathsf{p},t}$, χ_T is the penalty term, $\widehat{\boldsymbol{\eta}}_{\mathsf{p},t}$ is the value of $\widehat{\boldsymbol{\eta}}_t$ by letting the lag be p, and P is a sufficiently large fixed positive integer. The following theorem summaries the asymptotic property of (2.5).

Theorem 2.2. *Let Assumptions 1-3 hold. Suppose $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$, $\max_{t\ge 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s., $\chi_T \to 0$, and $c_T^{-2}\chi_T \to \infty$, where $c_T = h^2 + \left(\frac{\log T}{Th}\right)^{1/2}$. Then $\Pr(\widehat{\mathsf{p}} = p) \to 1$.*

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

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

Up to this point, we have estimated all unknown quantities of model (2.1).

2.2 Inference on Parameter Stability

Before moving on to investigate the impulse responses, we consider a hypothesis test which in practice is able to provide numerical evidence to justify the necessity of the model (2.1). An intuitive question to ask is whether some (if not all) components of the coefficient matrices are time–invariant. Formally, we may write it as follows:

$$\mathbb{H}_0 : \boldsymbol{C}\boldsymbol{\beta}(\cdot) = \boldsymbol{c} \text{ for some unknown } \boldsymbol{c} \in \mathbb{R}^s, \tag{2.6}$$

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

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

$$\widehat{Q}_{\boldsymbol{C},\boldsymbol{H}} = \int_0^1 \left\{\boldsymbol{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{c}}\right\}^\top \boldsymbol{H}(\tau) \left\{\boldsymbol{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{c}}\right\} \mathrm{d}\tau, \tag{2.7}$$

where $\widehat{\boldsymbol{\beta}}(\cdot) := \text{vec}[\widehat{\boldsymbol{A}}(\cdot)]$ should be obvious, and $\widehat{\boldsymbol{c}} = \int_0^1 \boldsymbol{C}\widehat{\boldsymbol{\beta}}(\tau)\mathrm{d}\tau$ is the estimator of $\boldsymbol{c}$. In (2.7), $\boldsymbol{H}(\cdot)$ is an $s \times s$ positive definite weighting matrix, and is typically set as the precision matrix associated with $\widehat{\boldsymbol{\beta}}(\cdot)$.

We now start to present the asymptotic properties of the proposed test. First, we present the asymptotic distribution of the test statistic.

Theorem 2.3. *Let Assumptions 1-3 hold. Suppose further that* $\max_{t\geq 1} E[\|\boldsymbol{\epsilon}_t\|^4|\mathcal{F}_{t-1}] < \infty$ *a.s.,* $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$*, each element of* $\boldsymbol{A}(\cdot)$ *has finite third-order derivative,* $Th^2/(\log T)^2 \to \infty$*, and* $Th^6 \to 0$*. If* $Th^{5.5} \to 0$ *and* $E[\|\boldsymbol{\epsilon}_t\|^\delta|\mathcal{F}_{t-1}] < \infty$ *a.s., then we have under* $\mathbb{H}_0$

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

where s *is the length of* $\boldsymbol{c}$*,* $\widehat{\boldsymbol{H}}(\tau) = (\boldsymbol{C}\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}(\tau)\boldsymbol{C}^\top)^{-1}$*,* $\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\Sigma}}^{-1}(\tau) \otimes \widehat{\boldsymbol{\Omega}}(\tau)$*,* $C_B = \int_0^2 (\int_{-1}^{1-v} K(u)K(u+v)\mathrm{d}u)^2\mathrm{d}v$*, and* $\widehat{\boldsymbol{\Sigma}}(\tau)$ *is defined in* (A.5).

Theorem 2.3 states that the test statistic converges to a normal distribution. The bias term $s\widetilde{v}_0$ can easily be calculated, and it arises due to the quadratic form of the test statistic. Here, we would like to emphasize that the proposed test is in fact a one–side test. Due to the quadratic form of (2.7), any departure from the true value will eventually yield a squared term when analysing the asymptotic power. Therefore, the null of (2.6) will be rejected at the level α if

$$\widehat{Q}^*_{\boldsymbol{C},\widehat{\boldsymbol{H}}} = \frac{T\sqrt{h}\left(\widehat{Q}_{\boldsymbol{C},\widehat{\boldsymbol{H}}} - \frac{1}{Th}s\widetilde{v}_0\right)}{\sqrt{4sC_B}} > q_{1-\alpha}, \tag{2.8}$$

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:

$$\mathbb{H}_1 : \boldsymbol{C}\boldsymbol{\beta}(\tau) = \boldsymbol{c} + d_T\boldsymbol{f}(\tau), \tag{2.9}$$

where $\boldsymbol{f}(\tau)$ is a twice continuously differentiable vector of functions, and $d_T \to 0$. The term $d_T\boldsymbol{f}(\tau)$ characterizes the departure of the time–varying coefficient $\boldsymbol{C}\boldsymbol{\beta}(\tau)$ from the constant *c*. Using the development of Theorem 2.3, it is straightforward to obtain the following corollary.

Corollary 2.1. *Let the conditions of Theorem 2.3 hold. Under the* $\mathbb{H}_1$ *of* (2.9)*, if* $d_T = T^{-1/2}h^{-1/4}$*, then*

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

where $\delta_1 = \int_0^1 \boldsymbol{f}(\tau)^\top(\boldsymbol{C}\boldsymbol{V}_{\boldsymbol{\beta}}(\tau)\boldsymbol{C}^\top)^{-1}\boldsymbol{f}(\tau)\mathrm{d}\tau$*. Moreover,* $\Pr\left(\widehat{Q}^*_{\boldsymbol{C},\widehat{\boldsymbol{H}}} > q_{1-\alpha}\right) \to \Phi\left(q_\alpha + \frac{\delta_1}{2\sqrt{sC_B}}\right)$.

Corollary 2.1 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., $\Pr(\widehat{Q}^*_{\boldsymbol{C},\widehat{\boldsymbol{H}}} > q_{1-\alpha}) \to 1$.

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 $Th^{5.5} \to 0$, are less restrictive than those assumed in the relevant literature, see, for example, $Th^{3.5} \to 0$ 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.

3 On Impulse Responses

In this section, we consider the estimation and inference of impulse responses. We first study the impulse response using short–run timing and long–run restrictions in Section 3.1, as both approaches do not require extra conditions. The economic interpretations of the two types of identification conditions can be found in Kilian and Lütkepohl (2017), so we do not repeat them below for the sake of space. In Section 3.2, we consider the use of external instruments, which has attracted some attentions recently (e.g., Stock and Watson, 2018; Paul, 2020).

3.1 SVAR Identification

As $\boldsymbol{\Omega}(\cdot) = \boldsymbol{\omega}(\cdot)\boldsymbol{\omega}^\top(\cdot)$, 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.

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 $\widehat{\boldsymbol{\Omega}}(\tau)$, i.e.,

$\widehat{\boldsymbol{\Omega}}(\tau) = \widehat{\boldsymbol{\omega}}(\tau)\widehat{\boldsymbol{\omega}}^\top(\tau)$. Alternatively, one can impose the conditions on the long–run impacts of the shocks (i.e., $\boldsymbol{B}(\tau)$ defined below). Specifically, define

$$\begin{aligned}\boldsymbol{B}(\tau) &:= \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau) = \left(\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i(\tau)\right)^{-1} \boldsymbol{\omega}(\tau),\\ \boldsymbol{\Psi}(\tau) &:= \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j(\tau) = \left(\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i(\tau)\right)^{-1},\end{aligned} \tag{3.1}$$

where $\boldsymbol{B}_j(\cdot)$ and $\boldsymbol{\Psi}_i(\cdot)$ are defined in Proposition 2.1, and the last equalities of both lines follow in an obvious matter. Thus, the elements of $\boldsymbol{B}(\tau)$ may be recovered from $\boldsymbol{B}(\tau)\boldsymbol{B}^\top(\tau) = \boldsymbol{\Psi}(\tau)\boldsymbol{\Omega}(\tau)\boldsymbol{\Psi}^\top(\tau)$. It is then convenient to assume that $\boldsymbol{B}(\tau)$ is a lower–triangular matrix, so $\widehat{\boldsymbol{B}}(\tau)$ is obtained from the Cholesky decomposition of $\widehat{\boldsymbol{\Psi}}(\tau)\widehat{\boldsymbol{\Omega}}(\tau)\widehat{\boldsymbol{\Psi}}^\top(\tau)$, where $\widehat{\boldsymbol{\Psi}}(\tau)$ is defined in the same way as $\boldsymbol{\Psi}(\tau)$ but replacing $\boldsymbol{A}_i(\tau)$ with $\widehat{\boldsymbol{A}}_i(\tau)$. Under the long–run restrictions, $\widehat{\boldsymbol{\omega}}(\tau) = \widehat{\boldsymbol{\Psi}}^{-1}(\tau)\widehat{\boldsymbol{B}}(\tau)$.

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

$$\widehat{\boldsymbol{B}}_j(\tau) = \widehat{\boldsymbol{\Psi}}_j(\tau)\widehat{\boldsymbol{\omega}}(\tau), \tag{3.2}$$

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

Theorem 3.1. *Under the conditions of Theorem 2.1. For any fixed integer $j \geq 0$*

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

where $\boldsymbol{B}_j^{(2)}(\tau) = \boldsymbol{C}_{j,1}(\tau)\mathrm{vec}\left(\boldsymbol{A}^{(2)}(\tau)\right) + \boldsymbol{C}_{j,2}(\tau)\mathrm{vech}\left(\boldsymbol{\Omega}^{(2)}(\tau)\right)$, and
$\boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau) = \left[\boldsymbol{C}_{j,1}(\tau), \boldsymbol{C}_{j,2}(\tau)\right]\boldsymbol{V}(\tau)\left[\boldsymbol{C}_{j,1}(\tau), \boldsymbol{C}_{j,2}(\tau)\right]^\top$. Specifically, under different identification conditions, we have the following expressions:

1. *Under the short–run timing restrictions,*

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

in which $\boldsymbol{N}_1(\tau) = (\boldsymbol{I}_{d^2} + \boldsymbol{K}_{d,d})(\boldsymbol{\omega}(\tau)\otimes\boldsymbol{I}_d)$, the elimination matrix $\boldsymbol{L}_d$ satisfies that

$\text{vech}(\boldsymbol{F}) = \boldsymbol{L}_d\text{vec}(\boldsymbol{F})$ *for any* $d \times d$ *matrix* $\boldsymbol{F}$*, and the commutation matrix* $\boldsymbol{K}_{m,n}$ *satisfies* $\boldsymbol{K}_{m,n}\text{vec}(\boldsymbol{G}) = \text{vec}(\boldsymbol{G}^\top)$ *for any* $m \times n$ *matrix* $\boldsymbol{G}$*.*

2. *Under the long–run restrictions,*

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

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

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

3.2 Identification with External Instruments

In addition to SVAR identification schemes, existing methods using external instruments for SVAR identification (SVAR–IV) are popular in the recent literature of macroeconomics (e.g., Stock and Watson, 2018; Paul, 2020). In this subsection, we would like to contribute along this line of research by specifically considering the case of Plagborg-Møller and Wolf (2021, p. 971). Formally, we are interested in the impulse responses of $\boldsymbol{x}_t$ to the structural shock $\epsilon_{1,t}$ (i.e., $\boldsymbol{\omega}_{\cdot,1}(\tau)$, the first column of $\boldsymbol{\omega}(\tau)$), and introduce an instrumental variable, π_t, which satisfies the following assumption.

Assumption 4. *Suppose* π_t *can be represented as* $\pi_t = \alpha_\pi(\tau_t)\epsilon_{1,t} + \sum_{j=1}^q \boldsymbol{\beta}_{j,\pi}^\top(\tau_t)\boldsymbol{x}_{t-j} + e_t$*, where* $\{e_t\}$ *is a sequence of independent variables with* $\max_t E|e_t|^\delta < \infty$ *and is independent of* $\{\boldsymbol{\epsilon_t}\}$*, and* q *is a nonnegative integer. In addition, each element of* $\alpha_\pi(\tau)$ *and* $\{\boldsymbol{\beta}_{j,\pi}(\tau)\}$ *is second order continuously differentiable on* $[0, 1]$*.*

Assumption 4 slightly extends the setting of Plagborg-Møller and Wolf (2021, p. 971) by including the lags of $\boldsymbol{x}_t$. Using Assumption 4, simple algebra shows that

$$\frac{E(\eta_{i,t}\pi_t)}{E(\eta_{1,t}\pi_t)} = \frac{\omega_{i,1}(\tau_t)}{\omega_{1,1}(\tau_t)} \quad \text{for } i = 2, \dots, d, \tag{3.3}$$

where $\eta_{i,t}$ stands for the i^{th} element of η_t, and $\omega_{i,j}(\cdot)$ denotes the $(i,j)^{th}$ element of $\boldsymbol{\omega}(\cdot)$. Thus, we rewrite the model (2.1) as

$$\boldsymbol{x}_t = \boldsymbol{\omega}^*_{\bullet,1}(\tau_t)x_{1,t} + \boldsymbol{A}^*(\tau_t)\boldsymbol{z}_{t-1} + \boldsymbol{\eta}^*_t, \tag{3.4}$$

where $\boldsymbol{\omega}^*_{\bullet,1}(\tau) = \frac{1}{\omega_{1,1}(\tau)}\boldsymbol{\omega}_{\bullet,1}(\tau)$, $\boldsymbol{A}^*(\tau) = \boldsymbol{A}(\tau) - \frac{1}{\omega_{1,1}(\tau)}\boldsymbol{\omega}_{\bullet,1}(\tau)\boldsymbol{A}_{1,\bullet}(\tau)$, and $\boldsymbol{\eta}^*_t = \sum_{j=2}^d[\boldsymbol{\omega}_{\bullet,j}(\tau_t) - \boldsymbol{\omega}_{\bullet,1}(\tau_t)\frac{\omega_{1,j}(\tau_t)}{\omega_{1,1}(\tau_t)}]\epsilon_{j,t}$. For model (3.4), we immediately obtain that

$$E(\boldsymbol{\eta}^*_t\pi_t) = 0. \tag{3.5}$$

Projecting out the the component $\boldsymbol{A}^*(\tau_t)\boldsymbol{z}_{t-1}$, a profile local linear IV estimator of $\boldsymbol{\omega}^*_{\bullet,1}(\tau)$ is then given by

$$\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) = [\boldsymbol{I}_d, \boldsymbol{0}_d]\left(\boldsymbol{W}_\tau^\top(\boldsymbol{I}_{dT} - \boldsymbol{S})^\top\boldsymbol{K}_\tau(\boldsymbol{I}_{dT} - \boldsymbol{S})\boldsymbol{X}_\tau\right)^{-1}\boldsymbol{W}_\tau^\top(\boldsymbol{I}_{dT} - \boldsymbol{S})^\top\boldsymbol{K}_\tau(\boldsymbol{I}_{dT} - \boldsymbol{S})\boldsymbol{x}, \tag{3.6}$$

where $\boldsymbol{S} = [\boldsymbol{s}(\tau_1)^\top\boldsymbol{Z}_0, \dots, \boldsymbol{s}(\tau_T)^\top\boldsymbol{Z}_{T-1}]^\top$, $\boldsymbol{s}(\tau) = [\boldsymbol{I}_{d^2p+d}, \boldsymbol{0}_{d^2p+d}](\boldsymbol{Z}_\tau^\top\boldsymbol{K}_\tau\boldsymbol{Z}_\tau)^{-1}\boldsymbol{Z}_\tau^\top\boldsymbol{K}_\tau$, $\boldsymbol{Z}_t = \boldsymbol{z}_t \otimes \boldsymbol{I}_d$, $\boldsymbol{K}_\tau = \mathrm{diag}(K_h(\tau_1 - \tau), \dots, K_h(\tau_T - \tau)) \otimes \boldsymbol{I}_d$, $\boldsymbol{x} = [\boldsymbol{x}_1^\top, \dots, \boldsymbol{x}_T^\top]^\top$,

$$\boldsymbol{Z}_\tau = \begin{bmatrix} \boldsymbol{Z}_0^\top & \boldsymbol{Z}_0^\top\frac{\tau_1-\tau}{h} \\ \vdots & \vdots \\ \boldsymbol{Z}_{T-1}^\top & \boldsymbol{Z}_{T-1}^\top\frac{\tau_T-\tau}{h} \end{bmatrix}, \quad \boldsymbol{W}_\tau = \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{W}_1\frac{\tau_1-\tau}{h} \\ \vdots & \vdots \\ \boldsymbol{W}_T & \boldsymbol{W}_T\frac{\tau_T-\tau}{h} \end{bmatrix}, \quad \boldsymbol{X}_\tau = \begin{bmatrix} \boldsymbol{X}_{1,1} & \boldsymbol{X}_{1,1}\frac{\tau_1-\tau}{h} \\ \vdots & \vdots \\ \boldsymbol{X}_{1,T} & \boldsymbol{X}_{1,T}\frac{\tau_T-\tau}{h} \end{bmatrix},$$

$\boldsymbol{W}_t = \pi_t \otimes \boldsymbol{I}_d$, and $\boldsymbol{X}_{1,t} = x_{1,t} \otimes \boldsymbol{I}_d$.

With $\widehat{\boldsymbol{\Psi}}_j(\tau)$ defined in (3.2) and $\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau)$ in (3.6), the SVAR–IV estimator of the impulse response function $\boldsymbol{B}_{j,\bullet1}(\tau)$ (i.e., the first column of $\boldsymbol{B}_j(\tau)$) for each $j \geq 0$ is given by

$$\widehat{\boldsymbol{B}}_{j,\bullet1}(\tau) = \widehat{\boldsymbol{\Psi}}_j(\tau)\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau). \tag{3.7}$$

We now establish the last theorem of this paper.

Theorem 3.2. *Let Assumptions 1–4 hold. Suppose that* $\max_{t\geq 1} E[\|\boldsymbol{\epsilon}_t\|^4|\mathcal{F}_{t-1}] < \infty$ *a.s.,* $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$ *and* $Th^5 \to \alpha \in [0,\infty)$*. Then, for* $\forall\tau \in (0,1)$ *and given* $j \geq 0$

$$\sqrt{Th}\left(\widehat{\boldsymbol{B}}_{j,\bullet1}(\tau) - \boldsymbol{B}_{j,\bullet1}(\tau) - \frac{1}{2}h^2\tilde{c}_2\boldsymbol{B}^{(2)}_{j,\bullet1}(\tau)\right) \to_D N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau)\right),$$

where

$$
\begin{aligned}
&\boldsymbol{B}^{(2)}_{j,\bullet 1}(\tau) = \boldsymbol{C}_{j,1}(\tau)\mathrm{vec}(\boldsymbol{A}^{(2)}(\tau)) + \boldsymbol{C}_{j,2}(\tau)\boldsymbol{\omega}^{*(2)}_{\bullet,1}(\tau),\\
&\boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau) = \boldsymbol{C}_{j,1}(\tau)\boldsymbol{V}_{1,1}(\tau)\boldsymbol{C}^\top_{j,1}(\tau) + \boldsymbol{C}_{j,2}(\tau)\boldsymbol{V}^*_{2,2}(\tau)\boldsymbol{C}^\top_{j,2}(\tau), \quad \boldsymbol{C}_{0,1}(\tau) = 0,\\
&\boldsymbol{C}_{j,1}(\tau) = \left(\boldsymbol{\omega}^{*\top}_{\bullet,1}(\tau)\otimes \boldsymbol{I}_d\right)\left(\sum_{m=0}^{j-1}\boldsymbol{J}(\boldsymbol{\Phi}^\top(\tau))^{j-1-m}\otimes \boldsymbol{\Psi}_m(\tau)\right)[\boldsymbol{0}_{d^2p\times d}, \boldsymbol{I}_{d^2p}], \ j\geq 1,\\
&\boldsymbol{C}_{j,2}(\tau) = \boldsymbol{\Psi}_j(\tau), \ j\geq 0.
\end{aligned}
$$

For the sake of space, we present the definitions of $\boldsymbol{V}_{1,1}$ and $\boldsymbol{V}^*_{2,2}$ in Appendix A. Similar to Theorem 3.1, $\boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau)$ can easily be estimated by replacing the unknown quantities with their estimates.

In what follows, we will examine the finite sample performance of the asymptotic properties of 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 (2.5), the bandwidth $\widehat{h}_{cv}$ is always chosen by minimizing the following cross-validation criterion function for each p.

$$
\widehat{h}_{cv} = \arg\min_h \sum_{t=1}^T \|\boldsymbol{x}_t - \widehat{\boldsymbol{a}}_{-t}(\tau_t) - \sum_{j=1}^{\mathsf{p}} \widehat{\boldsymbol{A}}_{j,-t}(\tau_t)\boldsymbol{x}_{t-j}\|^2, \tag{4.1}
$$

where $\widehat{\boldsymbol{a}}_{-t}(\cdot)$, and $\widehat{\boldsymbol{A}}_{j,-t}(\cdot)$ are obtained using (2.4) but leaving the t^{th} observation out. Once $\widehat{\mathsf{p}}$ and $\widehat{h}_{cv}$ are obtained, the estimation procedure is relatively straightforward.

We now 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 $\{\boldsymbol{x}_t\}$ to estimate the unrestricted model and the restricted model, and then compute $\widehat{Q}_{\boldsymbol{C},\widehat{\boldsymbol{H}}}$ based on (2.7).

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

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

Step 4: Repeat Steps 2-3 B times to obtain B bootstrap test statistics $\{\widetilde{Q}^b_{\boldsymbol{C},\widehat{\boldsymbol{H}}}\}_{b=1}^B$, as well as its empirical quantile $\widehat{q}_{1-\alpha}$. We reject the null hypothesis (2.6) at level α if $\widehat{Q}_{\boldsymbol{C},\widehat{\boldsymbol{H}}} > \widehat{q}_{1-\alpha}$.

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.

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

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

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

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

Next, we report the percentages of $\widehat{\mathsf{p}} < 2$, $\widehat{\mathsf{p}} = 2$, and $\widehat{\mathsf{p}} > 2$ respectively based on 1000 replications. Table 1 shows that the information criterion (2.5) performs reasonably well, as the percentages associated with $\widehat{\mathsf{p}} = 2$ are sufficiently close to 1 even for $T = 200$.

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

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

In addition, we evaluate the estimates of $\boldsymbol{A}(\tau)$, $\boldsymbol{\Omega}(\tau)$, as well as the estimates of the impulse responses (say, $\boldsymbol{B}_1(\tau)$ and $\boldsymbol{B}_5(\tau)$ 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:

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

where $\boldsymbol{\theta}(\cdot) \in \{\boldsymbol{A}(\cdot), \boldsymbol{\Omega}(\cdot), \boldsymbol{B}_1(\tau), \boldsymbol{B}_5(\tau)\}$, and $\widehat{\boldsymbol{\theta}}^{(n)}(\tau)$ 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 $\{\tau_t, t = 1, \ldots, T\}$, 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	$\boldsymbol{A}(\tau)$	$\boldsymbol{\Omega}(\tau)$	$\boldsymbol{B}_1(\tau)$	$\boldsymbol{B}_5(\tau)$
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)

Finally, we evaluate the estimates of the impulse responses $\{\boldsymbol{B}_{j,\cdot 1}\}$ based on the SVAR-IV method. We generate the instrument variable π_t by $\pi_t = \epsilon_{1,t} + e_t$, where $\{e_t\}$ is a sequence of i.i.d. standard normal variables. The coverage probability is calculated by the same way as in Table 2. As shown in Table 3, the empirical coverage probabilities are fairly close to 95% when $T = 800$ for all horizons.

Table 3: The empirical coverage probabilities of the SVAR-IV estimates of the impulse responses

T	$\boldsymbol{B}_{1,\cdot 1}(\tau)$	$\boldsymbol{B}_{3,\cdot 1}(\tau)$	$\boldsymbol{B}_{5,\cdot 1}(\tau)$	$\boldsymbol{B}_{7,\cdot 1}(\tau)$
200	0.869	0.897	0.945	0.939
400	0.910	0.923	0.953	0.942
800	0.926	0.928	0.951	0.941

4.3 Examining the Parameter Stability Test

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

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

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

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

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 $\boldsymbol{A}_1(\cdot)$ 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:

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

Table 4 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 the 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 Chapter 3 of Gao (2007), and 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.

Table 4: 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

5 Time–Varying Government Spending Multipliers

In this section, we investigate whether US government spending multipliers (i.e., the change in output caused by \$1 change in government spending) vary over time. The estimates of government spending multipliers are crucial to fiscal policy analysis since it measures to which extent government purchases can stimulate private activity. Along this line of research, one important question is that whether the US economy has changed over time so that estimates from historical data are unreliable for modern policy analysis. As Ramey and Zubairy (2018) put it, "*Theory tells us that details such as the persistence of spending changes, how they are financed, how monetary policy reacts, and the tightness of the labor* *market can significantly affect the magnitude of the multipliers. Unfortunately, the data do not present us with clean natural experiments that can answer these questions.*" Although the literature has begun to explore whether the estimates of multipliers depend on the economy states (e.g., Ramey and Zubairy, 2018; Barnichon et al., 2022), few studies aim to quantity the varying government spending multipliers over time. In what follows, we address this issue using the newly proposed 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 two commonly adopted macroeconomic variables of the literature (Blanchard and Perotti, 2002; Ramey and Zubairy, 2018), which are government spending and real per capita GDP, each divided by trend GDP.[2] In addition, following Ramey and Zubairy (2018) and Barnichon et al. (2022), we use the defense news variables scaled by trend GDP as the instrumental variable for identifying structural government spending shocks. For detailed descriptions of these three variables and the justification of the instrumental relevance, see Ramey and Zubairy (2018). The data are quarterly observations from 1954:Q1 to 2015:Q4, which are collected from supplementary document of Ramey and Zubairy (2018). Figure 1 plots the three variables.

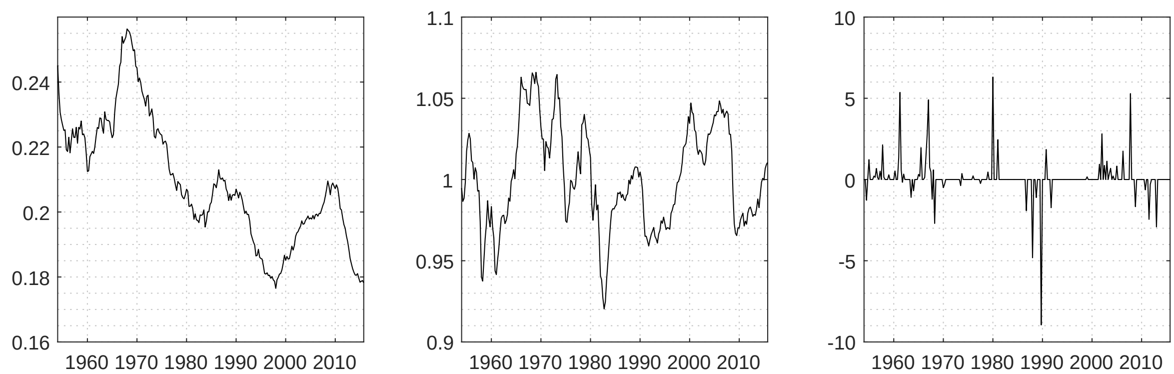

Figure 1: Plots of scaled government spending (left), scaled real per capita GDP (middle) and scaled military news (% of GDP, right)

For the time–varying VAR(p) model, the optimal lag is $\widehat{\mathsf{p}} = 2$ by our approach, while it is often assumed to be known with the value varying from 2 to 4 in the literature. Thus, the following analyses focus on the time–varying VAR(2) model (referred to as TV–VAR(2) hereafter). We then conduct robustness checks to see whether the maximum eigenvalue of the companion matrix is less than 1 and to see whether the innovation process $\boldsymbol{\eta}_t$ exhibits serial correlation. The estimation results suggest the value of $|\lambda_{\max}\{\widehat{\mathbf{\Phi}}(\tau)\}|$ varies from 0.7–0.9 over time, and thus Assumption 1.1 is automatically met. In addition, we use the

[2]Following Ramey and Zubairy (2018), the trend GDP is estimated as a sixth-degree polynomial for the logarithm of GDP.

multivariate version of Breusch–Godfrey LM test (Breusch, 1978; Godfrey, 1978) to test the autocorrelation of the reduce–form residuals $\boldsymbol{\eta}_t$'s. The null hypothesis is $H_0 : E\left(\boldsymbol{\eta}_t \boldsymbol{\eta}_{t-h}^\top\right) = 0$ for $h = 1, \ldots, 5$, while the p–values are 0.57, 0.44, 0.21, 0.16, 0.23 respectively suggesting that the TV–VAR(2) model fits the data quite well.

Before investigating the government spending multipliers, we further check whether the VAR coefficients (i.e., the policy transmission mechanism) are time–varying. We employ the proposed test statistic to examine the constancy of model coefficients, and summarize the results in Table 5. From Table 5 (the row labelled "Constancy"), we choose the TV–VAR(2) model over a constant VAR model. We then apply the testing procedure to distinguish time–varying intercept or time–varying autoregressive coefficients. By Table 5, at the 5% significance level we conclude that both the intercept term and VAR coefficients are time–varying[3], implying that there exists significant time–variations in policy transmission mechanism.

Table 5: Summary statistics of the test

	test statistic	p-value
Constancy	149.32	0.00
$\boldsymbol{a}(\cdot)$	35.26	0.00
$\boldsymbol{A}_1(\cdot), \boldsymbol{A}_2(\cdot)$	35.72	0.00

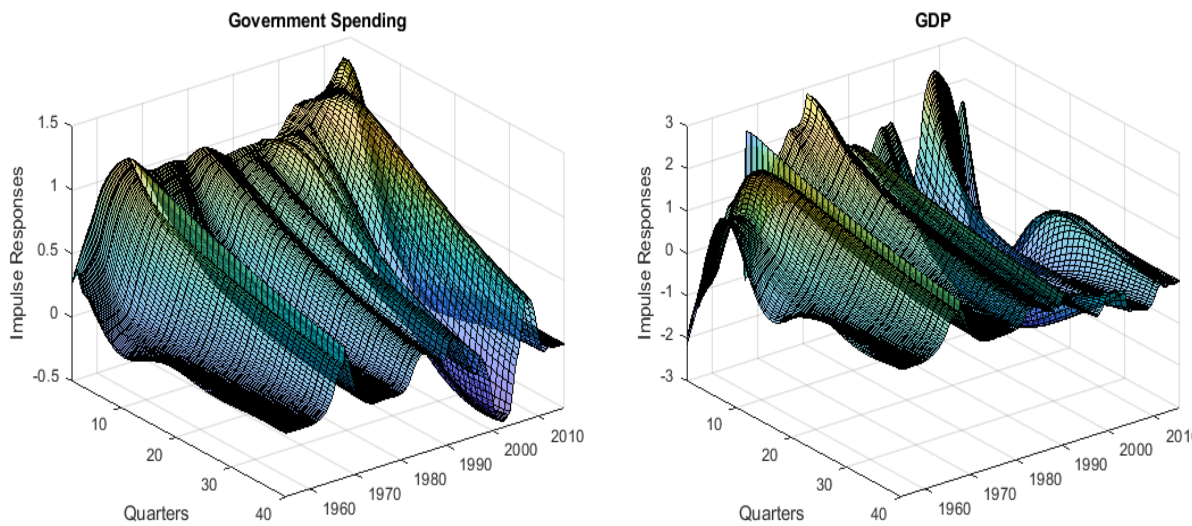

Figure 2: Time–varying impulse responses to a military news shock. Left front axis left: Quarters (Horizon). Right front axis left: Time.

We now discuss how government spending and GDP respond to a military news shock over time. Figure 2 plots the time–varying impulse responses of government spending and

[3]Certainly, 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.

GDP to a military news shock. The size of the shock is normalized to 1 percent of GDP. Clearly, these responses vary over time, especially for GDP, indicating a substantial time–variation in the policy transmission mechanism. Interestingly, our results show that the responses of GDP are negative at some horizons after 2000, while Ramey and Zubairy (2018) find some negative responses of GDP at high employment state.

We then calculate the time–varying government spending multipliers based on the cumulative responses of government spending and GDP. Following Ramey and Zubairy (2018), the multipliers are calculated as the integral of the GDP response divided by the integral government spending response. Figure 3 shows the time–varying government multipliers for a 2–year and 4–year horizon, as well as their 95% confidence intervals. The standard errors are calculated based on Theorem 3.2. From Figure 3, it can be seen that government multipliers are decreasing and are below unity after 1990s, while the government multipliers are quite stable and around 2 before that. Specifically, government multipliers are not significantly different from zeros after 1990s indicating US stimulus policy do not stimulate private activity in recent years. One possible explanation may be the arguement given by Ramey and Zubairy (2018), *"Increases over time in financial market access and consumer sophistication should reduce the fraction of rule-of-thumb consumers, thus reducing multipliers in recent years."* Importantly, our results provide a solid evidence for this argument.

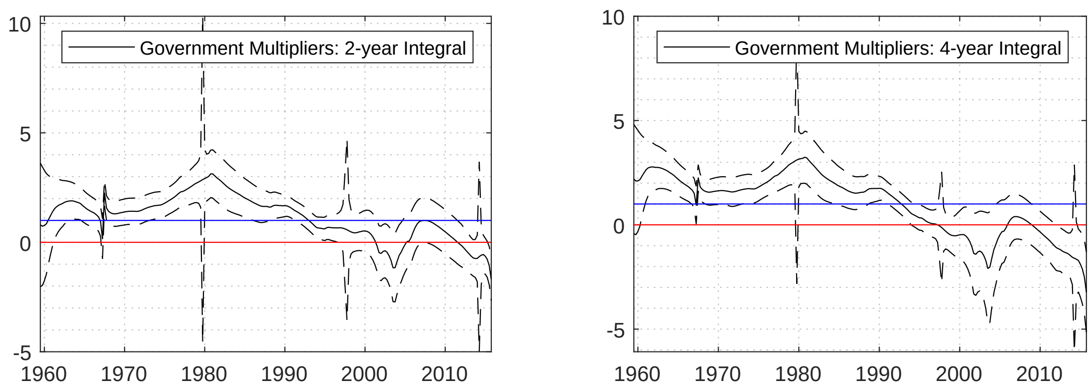

Figure 3: Estimates of time–varying government multipliers based on SVAR-IV approach for a 2-year and 4-year horizon, as well as their 95% condidence intervals.

Finally, we explore the sensitivity of our findings to different identification conditions. Alternatively, we adopt the short–run timing identification to estimate time–varying multipliers. This identification scheme is also adopted in Blanchard and Perotti (2002), Ramey and Zubairy (2018) and Barnichon et al. (2022), and is based on the assumption that within–quarter government spending does not contemporaneously respond to macroeconomic variables. Figure 4 plots the estimates of time–varying multipliers based on the short–run timing restrictions. Figure 4 reveals similar results that government multipliers decreasing over time and are not significantly different from zeros after 1990s.

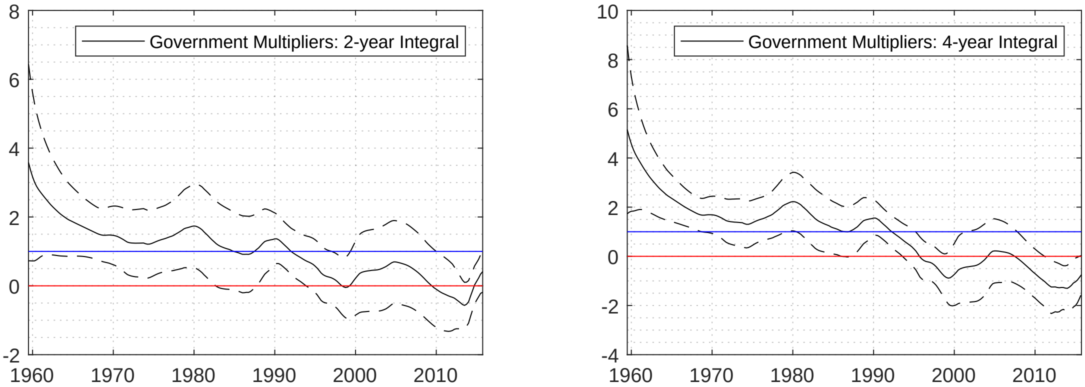

Figure 4: Estimates of time–varying government multipliers based on short-run timing restriction for a 2-year and 4-year horizon, as well as their 95% condidence intervals.

Overall, we find government spending multipliers that are above one before 1990s and are not significantly different from zero after 1990s across different identification schemes.

6 Conclusions and Discussion

In this paper, we have investigated a 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 an information criterion to select the optimal lag, an integrated L_2 type test to determine the constant coefficients, and the impulse responses analyses using SVAR identification schemes and then external instruments. Simulation studies are conducted to evaluate the theoretical findings. Finally, we demonstrate the empirical relevance of the proposed methods through an application to estimating time–varying government multipliers. We find government spending multipliers that are above one before 1990s and are not significantly different from zero after 1990s across different identification schemes.

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 summarize the relevant mathematical symbols in this appendix. The proofs of the theoretical results are provided in the online Appendix B of the paper.

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

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

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

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

where $\boldsymbol{z}_t$ is defined in (2.3).

Next, we let

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

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

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

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

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

where $\widehat{\boldsymbol{V}}_{1,1}(\tau)$, $\widehat{\boldsymbol{V}}_{2,1}(\tau)$ and $\widehat{\boldsymbol{V}}_{2,2}(\tau)$ 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 (2.4).

Recall the following definition:

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

Replacing $\boldsymbol{A}_j(\tau)$'s of (A.6) with their estimators obtained from (2.4) yields an estimator, $\widehat{\boldsymbol{\Phi}}(\tau)$, for $\boldsymbol{\Phi}(\tau)$.

Finally, we define

$$\boldsymbol{V}^*_{2,2}(\tau) = (\Sigma^{*,-2}(\tau)\Delta^*(\tau)) \otimes \boldsymbol{\Omega}^*(\tau), \tag{A.7}$$

where $\Delta^*(\tau) = \Sigma_\pi(\tau) - \boldsymbol{\Sigma}^\top_{\boldsymbol{z}\pi}(\tau)\boldsymbol{\Sigma}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{z}\pi}(\tau)$ and $\Sigma^*(\tau) = \Sigma_{\pi x_1}(\tau) - \boldsymbol{\Sigma}^\top_{\boldsymbol{z}\pi}(\tau)\boldsymbol{\Sigma}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{z}x_1}(\tau)$ with $\boldsymbol{\Omega}^*(\tau) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T E(\boldsymbol{\eta}^*_t \boldsymbol{\eta}^{*,\top}_t) K_h(\tau_t - \tau)$, $\boldsymbol{\Sigma}_{\boldsymbol{z}x_1} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T E(\boldsymbol{z}_{t-1} x_{1,t}) K_h(\tau_t - \tau)$, $\boldsymbol{\Sigma}_{\boldsymbol{z}\pi}(\tau) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T E(\boldsymbol{z}_{t-1}\pi_t) K_h(\tau_t - \tau)$ and $\Sigma_{\pi x_1}(\tau) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T E(\pi_t x_{1,t}) K_h(\tau_t - \tau)$.

References

Barnichon, R., Debortoli, D. and Matthes, C. (2022), 'Understanding the size of the government spending multiplier: It's in the sign', *Review of Economic Studies* **89**(1), 87–117.

Blanchard, O. and Perotti, R. (2002), 'An empirical characterization of the dynamic effects of changes in government spending and taxes on output', *Quarterly Journal of Economics* **117**(4), 1329–1368.

Breusch, T. S. (1978), 'Testing for autocorrelation in dynamic linear models', *Australian Economic Papers* **17**(31), 334–355.

Cai, Z. (2007), 'Trending time–varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**(2), 163–188.

Chen, B. and Hong, Y. (2012), 'Testing for smooth structural changes in time series models via nonparametric regression', *Econometrica* **80**(3), 1157–1183.

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.

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

Gao, J. (2007), *Nonlinear Time Series: Semi– and Non–Parametric Methods*, Chapman & Hall/CRC, London.

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

Gao, J. and Hawthorne, K. (2006), 'Semiparametric estimation and testing of the trend of temperature series', *Econometrics Journal* **9**(3), 387–408.

Giraitis, L., Kapetanios, G. and Yates, T. (2014), 'Inference on stochastic time-varying coefficient models', *Journal of Econometrics* **179**(1), 46–65.

Giraitis, L., Kapetanios, G. and Yates, T. (2018), 'Inference on multivariate heteroscedastic time varying random coefficient models', *Journal of Time Series Analysis* **39**(2), 129–149.

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.

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* **39**, Forthcoming.

Kapetanios, G., Marcellino, M. and Venditti, F. (2019), 'Large time-varying parameter vars: A nonparametric approach', *Journal of Applied Econometrics* **34**(7), 1027–1049.

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

Li, D., Phillips, P. C. B. and Gao, J. (2020), 'Kernel-based inference in time-varying coefficient cointegrating regression', *Journal of Econometrics* **215**(2), 607–632.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.

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

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

Plagborg-Møller, M. and Wolf, C. K. (2021), 'Local projections and vars estimate the same impulse responses', *Econometrica* **89**(2), 955–980.

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

Ramey, V. A. and Zubairy, S. (2018), 'Government spending multipliers in good times and in bad: evidence from us historical data', *Journal of Political Economy* **126**(2), 850–901.

Richter, S. and Dahlhaus, R. (2019), 'Cross validation for locally stationary processes', *Annals of Statistics* **47**(4), 2145–2173.

Robinson, P. M. (1989), 'Chapter 15: Nonparametric estimation of time-varying parameters', *Statistical Analysis and Forecasting of Economic Structural Change* pp. 253–264.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica* **48**(1), 1–48.

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.

Stock, J. H. and Watson, M. W. (2018), 'Identification and estimation of dynamic causal effects in macroeconomics using external instruments', *Economic Journal* **128**(610), 917–948.

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.

Xu, Z., Kim, S. and Zhao, Z. (2021), 'Locally stationary quantile regression for inflation and interest rates', *Journal of Business and Economic Statistics* **39**, Forthcoming.

Yan, Y., Gao, J. and Peng, B. (2021), Asymptotics for time-varying vector moving average (infinity) processes. Working paper at https://ideas.repec.org/p/msh/ebswps/2021-22.html.

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

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 presents the proofs of the main results of the paper, while Appendix B.3 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* $s_T^{-2}U_T^2 \to_P 1$ *and* $\sum_{t=1}^T E[Z_{T,t}^2 I(|Z_{T,t}| > \nu)] \to 0$ *for any* $\nu > 0$ *with* $Z_{T,t} = s_T^{-1} Z_t$*, then as* $T \to \infty$*,* $s_T^{-1} S_T \to_D N(0,1)$*.*

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*,* $t = 1, \ldots, T$*. Let* $V_T = \sum_{t=1}^T \mathrm{Var}\,(Z_t|\mathcal{F}_{t-1}) \leq V$ *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. $\max_t \|\boldsymbol{\mu}_t\| < \infty$, $\max_t \sum_{j=1}^\infty j\|\boldsymbol{B}_{j,t}\| < \infty$, $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \|\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\| < \infty$ *and* $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^\infty j\|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\| < \infty$*. Define the stochastic process of the form* $\boldsymbol{h}_t = \boldsymbol{\mu}_t + \sum_{j=0}^\infty \boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_t$ *for* $t = 1, ..., T$*.*

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* $\{\boldsymbol{W}_{T,t}(\cdot)\}_{t=1}^T$ *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. $\sup_{\tau\in[a,b]} \sum_{t=1}^T \|\boldsymbol{W}_{T,t}(\tau)\| = O(1)$*;*
2. $\sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\| = O(d_T)$*, where* $d_T = \sup_{\tau\in[a,b],t\geq 1} \|\boldsymbol{W}_{T,t}(\tau)\|$*.*

Then as $T \to \infty$*,*

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

Lemma B.3 is the Lemma 2.3 in Yan et al. (2021).

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

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

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

Lemma B.4 is the Lemma 2.2 in Yan et al. (2021).

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

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

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

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

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

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

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

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

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

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

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

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

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

1. *if* $\mathsf{p}\geq p$*, then* $RSS(\mathsf{p})=\frac{1}{T}\sum_{t=1}^T E\left(\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t\right)+O_P\left(\rho_T^2\right)$ *with* $\rho_T=h^2+\sqrt{\frac{\log(T)}{Th}}$*;*

2. *if* $\mathsf{p}<p$*, then* $RSS(\mathsf{p})=\frac{1}{T}\sum_{t=1}^T E\left(\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t\right)+c+o_P(1)$ *with some constant* $c>0$*.*

Lemma B.8. *Let Assumptions 1–4 hold. Suppose* $\frac{T^{1-\frac{4}{\delta}}h}{\log T}\to\infty$ *and* $\max_{t\geq 1}E[\|\boldsymbol{\epsilon}_t\|^4|\mathcal{F}_{t-1}]<\infty$ *a.s. Then,*

1. $\sup_{\tau\in[h,1-h]}\left\|\boldsymbol{s}(\tau)\boldsymbol{X}_\tau-\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{X}_1}(\tau)\otimes[1,0]\right\| = O_P(\rho_T)$*, where* $\boldsymbol{\Sigma}_{\boldsymbol{Z}}(\tau)=\boldsymbol{\Sigma}(\tau)\otimes\boldsymbol{I}_d$*,* $\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{X}_1}(\tau)=\boldsymbol{\Sigma}_{\boldsymbol{z}x_1}\otimes\boldsymbol{I}_d$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{z}x_1}=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\boldsymbol{z}_{t-1}x_{1,t})K_h(\tau_t-\tau)$*;*

2. $\sup_{\tau\in[h,1-h]}\left\|\boldsymbol{s}(\tau)\boldsymbol{W}_\tau-\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{W}}(\tau)\otimes[1,0]\right\|=O_P(\rho_T)$*, where* $\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{W}}(\tau)=\boldsymbol{\Sigma}_{\boldsymbol{z}\pi}(\tau)\otimes\boldsymbol{I}_d$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{z}\pi}(\tau)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\boldsymbol{z}_{t-1}\pi_t)K_h(\tau_t-\tau)$*;*

3. $\sup_{\tau\in[h,1-h]}\left\|\frac{1}{T}\boldsymbol{W}_\tau^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top\boldsymbol{K}_\tau(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{X}_\tau-(\Sigma^*(\tau)\otimes\boldsymbol{I}_d)\otimes\boldsymbol{\Lambda}_1\right\|=O_P(\rho_T)$*, where* $\Sigma^*(\tau)=\Sigma_{\pi x_1}(\tau)-\boldsymbol{\Sigma}_{\boldsymbol{z}\pi}^\top(\tau)\boldsymbol{\Sigma}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{z}x_1}(\tau)$*,* $\Sigma_{\pi x_1}(\tau)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\pi_t x_{1,t})K_h(\tau_t-\tau)$ *and* $\boldsymbol{\Lambda}_1=\mathrm{diag}(1,\widetilde{c}_2)$*;*

4. *if* $Th^9\to 0$*,*

$$\boldsymbol{W}_\tau^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top\boldsymbol{K}_\tau(\boldsymbol{I}_{dT}-\boldsymbol{S})\begin{bmatrix}\boldsymbol{Z}_0^\top\mathrm{vec}(\boldsymbol{A}^*(\tau_1))\\ \vdots\\ \boldsymbol{Z}_{T-1}^\top\mathrm{vec}(\boldsymbol{A}^*(\tau_T))\end{bmatrix}=o_P(\sqrt{Th});$$

5. *if* $Th^9\to 0$*,*

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^T[\boldsymbol{W}_t-\ \boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{W}}^\top(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t)\boldsymbol{Z}_{t-1}(1+O_P(\rho_T))]$$
$$\times\boldsymbol{Z}_{t-1}^\top\boldsymbol{s}(\tau_t)\begin{bmatrix}\boldsymbol{X}_{1,1}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_1)-\boldsymbol{\omega}_{\cdot,1}^*(\tau_t))\\ \vdots\\ \boldsymbol{X}_{1,T}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_T)-\boldsymbol{\omega}_{\cdot,1}^*(\tau_t))\end{bmatrix}K_h(\tau_t-\tau)=o_P(1).$$

Define $w_{s,t}=\frac{1}{T\sqrt{h}}\int_{-1}^1 K(u)K\left(u+\frac{t-s}{Th}\right)\mathrm{d}u$. 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.9. *Suppose Assumption 3 hold. Then, we have*

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

2. $\max_{2\le t\le T} a_t = O(1/T)$;

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

Lemma B.10. *Let $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_T$ be a d-dimensional martingale difference for which $\boldsymbol{\xi}_t \in \mathcal{L}^p$, $p > 1$. Let $p^* = \min(2, p)$. Then*

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

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

Lemma B.11. *Suppose Assumption 3 hold. Assume $\boldsymbol{w}_t^* \in \mathcal{L}^{\delta/2}$ for some $\delta > 4$ is $\mathcal{F}_t$-measurable. Then, as $T \to \infty$,*

$$E\left|\sum_{t=2}^{T}\mathrm{tr}\left[\left(\boldsymbol{w}_t^* - E(\boldsymbol{w}_t^*)\right)\boldsymbol{h}_{t-1}^*\boldsymbol{h}_{t-1}^{*,\top}\right]\right| \to 0.$$

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

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

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

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

Lemma B.13. *Suppose Assumptions 1–3 hold and $\boldsymbol{y}_t = \boldsymbol{Z}_{t-1}\boldsymbol{\eta}_t$. Then as $T \to \infty$,*

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

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

Lemma B.14. *Let Assumptions 1-3 hold. Suppose further that $\max_{t\ge1} E[\|\boldsymbol{\epsilon}_t\|^4|\mathcal{F}_{t-1}] < \infty$ a.s., $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$, each element of $\boldsymbol{A}(\cdot)$ has finite third-order derivative, $Th^2/(\log T)^2 \to \infty$, and $Th^6 \to 0$. As $T \to \infty$, we have*

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

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

B.2 Proofs of the Main Results

Proof of Proposition 2.1.

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

First, we investigate the validity of the VMA representations of $\boldsymbol{x}_t$ and $\widetilde{\boldsymbol{x}}_t$. Let ρ_A denote the largest eigenvalue of $\boldsymbol{\Phi}(\tau)$ uniformly over $\tau \in [0,1]$. Then, $\rho_A < 1$ by Assumption 1.1. Similar to the proof of Proposition 2.4 in Dahlhaus and Polonik (2009), we have $\max_{t\geq 1}\left\|\prod_{m=0}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\right\| \leq M\rho_A^j$. It follows that $\|\mathrm{E}(\boldsymbol{x}_t)\| \leq \sum_{j=0}^{\infty}\|\boldsymbol{\Psi}_{j,t}\| \cdot \|\boldsymbol{a}(\tau_{t-j})\| \leq M\sum_{j=0}^{\infty}\rho_A^j < \infty$ and

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

Similarly, we have $\|\mathrm{E}(\widetilde{\boldsymbol{x}}_t)\| < \infty$ and $\|\mathrm{Var}(\widetilde{\boldsymbol{x}}_t)\| < \infty$.

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

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

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

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

The proof is now completed. □

Proof of Theorem 2.1.

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

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

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

We now begin our investigation. Since

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

we write

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

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 $I_{T,3} = O_P\left((\frac{\log T}{Th})^{\frac{1}{2}}\right)$ uniformly over $\tau \in [0,1]$. Therefore, the first result follows.

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

$$
\begin{aligned}
\widehat{\boldsymbol{\Omega}}(\tau) &= \frac{1}{Th}\sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) + O_P\left(\frac{1}{Th}\right) \\
&= \frac{1}{Th}\sum_{t=1}^T(\boldsymbol{\eta}_t + \widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\boldsymbol{\eta}_t + \widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) + O_P\left(\frac{1}{Th}\right) \\
&= \frac{1}{Th}\sum_{t=1}^T \boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) + \frac{1}{Th}\sum_{t=1}^T(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) \\
&\quad + \frac{1}{Th}\sum_{t=1}^T \boldsymbol{\eta}_t(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) + \frac{1}{Th}\sum_{t=1}^T(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)\boldsymbol{\eta}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) + O_P\left(\frac{1}{Th}\right) \\
&:= \boldsymbol{I}_{T,4} + \boldsymbol{I}_{T,5} + \boldsymbol{I}_{T,6} + \boldsymbol{I}_{T,7} + O_P\left(\frac{1}{Th}\right).
\end{aligned}
$$

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

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

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

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

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

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

Below, we focus on $\boldsymbol{I}_{T,8}$. First, we show $\text{Var}(\boldsymbol{I}_{T,8}) \to \boldsymbol{V}(\tau)$. Let $\text{Var}(\boldsymbol{I}_{T,8}) = \begin{bmatrix} \widetilde{\boldsymbol{V}}_{1,1}(\tau) & \widetilde{\boldsymbol{V}}_{2,1}^\top(\tau) \\ \widetilde{\boldsymbol{V}}_{2,1}(\tau) & \widetilde{\boldsymbol{V}}_{2,2}(\tau) \end{bmatrix}$, where the definition of each block should be obvious. Moreover, simple algebra shows that $\widetilde{\boldsymbol{V}}_{i,j}(\tau) \to \boldsymbol{V}_{i,j}(\tau)$ for $i,j \in \{1,2\}$.

By construction and Assumption 2, $\boldsymbol{I}_{T,8}$ 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 $\boldsymbol{d}^\top \boldsymbol{I}_{T,8} \to_D N\left(\boldsymbol{0}, \boldsymbol{d}^\top \boldsymbol{V}(\tau)\boldsymbol{d}\right)$ for any conformable unit vector $\boldsymbol{d}$. Let

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

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

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

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 $\sum_{t=1}^{T} E(\boldsymbol{Z}_{T,t}^2(\tau)|\mathcal{F}_{t-1}) \to_P \boldsymbol{d}^\top \boldsymbol{V}(\tau)\boldsymbol{d}$.

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

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

Thus, the CLT follows.

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

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

and

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

The proof is now completed. □

Proof of Theorem 2.2.

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

Note that

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

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

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

$$\Pr\left(\text{IC}(\mathsf{p}) < \text{IC}(p)\right) \leq \Pr\left(\log[\text{RSS}(\mathsf{p})/\text{RSS}(p)] + \chi_T < 0\right) \to 0.$$

The proof is now completed. □

Proof of Theorem 2.3.

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

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

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

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

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

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

$$\begin{aligned}
&\sup_{[h,1-h]} \|\boldsymbol{S}_T(\tau) - \mathbf{\Sigma}_{\boldsymbol{Z}}(\tau)\otimes\mathbf{\Lambda}_1\| = O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right),\\
&\sup_{[0,1]} \|\boldsymbol{R}_T(\tau)\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right),
\end{aligned} \tag{B.3}$$

where $\mathbf{\Sigma}_{\boldsymbol{Z}}(\tau) = \mathbf{\Sigma}(\tau)\otimes\boldsymbol{I}_d$, $\mathbf{\Lambda}_1 = \mathrm{diag}(\widetilde{c}_0,\widetilde{c}_2)$ and $\boldsymbol{R}_T(\tau) = \frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_{t-1}\boldsymbol{\eta}_t K_h(\tau_t-\tau)$. Hence, under the null hypothesis, we have

$$\sup_{[h,1-h]} \left\|\boldsymbol{C}\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{c} - \boldsymbol{C}\mathbf{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau)\boldsymbol{R}_T(\tau)\right\| = O_P(\rho_T^2). \tag{B.4}$$

By (B.3) and $Th^{11/2}\to 0$, we have

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

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

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

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

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

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

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

Then we have

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

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

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

we have

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

where the definition of $\widetilde{U}$ is obvious. By Lemma B.13, we have

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

where $\sigma_{\widetilde{U}}^2=\lim_{T\to\infty}\sum_{t=2}^T\mathrm{tr}\left\{E\left(\boldsymbol{H}_0^\top(\tau_t)\boldsymbol{Z}_{t-1}\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top\boldsymbol{Z}_{t-1}^\top\boldsymbol{H}_0(\tau_t)\right)E\left(\sum_{s=1}^{t-1}\boldsymbol{Z}_{s-1}\boldsymbol{\eta}_s\boldsymbol{\eta}_s^\top\boldsymbol{Z}_{s-1}^\top\right)w_{s,t}^2\right\}$. We then show that $\sigma_{\widetilde{U}}^2 = C_B\mathrm{tr}\left\{\int_0^1\boldsymbol{\Sigma}_Q(\tau)^2\mathrm{d}\tau\right\}$. Let $\boldsymbol{V}_1(\tau)=\boldsymbol{H}_0(\tau)\boldsymbol{V}_2(\tau)\boldsymbol{H}_0(\tau)$ and $\boldsymbol{V}_2(\tau)=\boldsymbol{\Sigma}(\tau)\otimes\boldsymbol{\Omega}(\tau)$. Write

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

$$
\begin{aligned}
&= \frac{1}{T^2 h}\sum_{s=1}^{T-1}\sum_{j=1}^{T-s}\boldsymbol{V}_1(\tau_s)\boldsymbol{V}_2(\tau_s)\left[\int_{-1}^{1}K\left(u\right)K\left(u+\frac{j}{Th}\right)\mathrm{d}u\right]^2 \\
&\quad + \frac{1}{T^2 h}\sum_{s=1}^{T-1}\sum_{j=1}^{T-s}O(j/T)\boldsymbol{V}_2(\tau_s)\left[\int_{-1}^{1}K\left(u\right)K\left(u+\frac{j}{Th}\right)\mathrm{d}u\right]^2 := I_{T,3}+I_{T,4},
\end{aligned}
$$

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

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

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

Combining the above results, we have proved

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

Note that

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

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

Since $\widehat{\boldsymbol{c}} = \boldsymbol{c} + O_P\left(T^{-1/2}\right)$ by Lemma B.14, we have $I_{T,5} = O_P(T^{-1})$. For $I_{T,6}$, by (B.1) and (B.4), we have

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

provided that

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

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

Observe that

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

Finally, we need only to focus on

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

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

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

The proof is now completed. □

Proof of Corollary 2.1.

Under the local alternative (2.9), we have $\boldsymbol{C}\boldsymbol{\beta}(\tau)=\boldsymbol{c}+d_T\boldsymbol{f}(\tau)$ and thus

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

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

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

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

For $I_{T,2}$, since $\sup_{\tau\in[0,1]}\|\boldsymbol{R}_T(\tau)\|=O_P\left(\sqrt{\frac{\log T}{Th}}\right)$, 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 $T\sqrt{h}\left(\int_0^1 \boldsymbol{R}_T^\top(\tau)\boldsymbol{H}_0(\tau)\boldsymbol{R}_T(\tau) - (Th)^{-1}A_Q\right) \to N(0, B_Q)$, we have

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

In addition, similar to the proof of Theorem 2.3, we have $T\sqrt{h}\left(\widehat{Q}_{\boldsymbol{C},\boldsymbol{H}} - \widehat{Q}_{\boldsymbol{C},\widehat{\boldsymbol{H}}}\right) = o_P(1)$. The proof is now completed. $\square$

Proof of Theorem 3.1.

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

Let $\boldsymbol{\alpha}(\tau) = \text{vec}\left(\boldsymbol{A}_1(\tau), ..., \boldsymbol{A}_p(\tau)\right)$, $\boldsymbol{\sigma}(\tau) = \text{vech}\left(\boldsymbol{\Omega}(\tau)\right)$ and $\boldsymbol{\phi}(\tau) = [\boldsymbol{\alpha}^\top(\tau), \boldsymbol{\sigma}^\top(\tau)]^\top$. Given the joint distribution of $\text{vec}(\widehat{\boldsymbol{A}}(\tau))$ and $\text{vech}(\widehat{\boldsymbol{\Omega}}(\tau))$ in Theorem 2.1, Theorem 3.1 can be obtained by the Delta method. By the first-order approximation of $\text{vec}\left(\widehat{\boldsymbol{B}}_j(\tau)\right)$ around $\text{vec}\left(\boldsymbol{B}_j(\tau)\right)$, we have

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

and thus

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

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

To complete the proof, we first derive an analytic form for the derivative $\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{B}_j(\tau)$ 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 $\text{d}\boldsymbol{\Omega}(\tau) = \text{d}\boldsymbol{\omega}(\tau)\cdot\boldsymbol{\omega}^\top(\tau) + \boldsymbol{\omega}(\tau)\cdot\text{d}\boldsymbol{\omega}^\top(\tau)$. Let $\boldsymbol{B}$ and $\boldsymbol{C}$ be $n\times q$ and $q\times r$ matrices, respectively. By $\text{vec}(\boldsymbol{ABC}) = \boldsymbol{C}^\top\otimes\boldsymbol{A}\text{vec}(\boldsymbol{B})$, $\text{vec}(\boldsymbol{A}^\top) = \boldsymbol{K}_{m,n}\text{vec}(\boldsymbol{A})$ and $\boldsymbol{K}_{m,q}(\boldsymbol{A}\otimes\boldsymbol{C}) = (\boldsymbol{C}\otimes\boldsymbol{A})\boldsymbol{K}_{n,r}$, we have $\boldsymbol{N}_1(\tau)\text{vec}(\text{d}\boldsymbol{\omega}(\tau)) = \text{vec}(\text{d}\boldsymbol{\Omega}(\tau))$, where $\boldsymbol{N}_1(\tau) = (\boldsymbol{I}_{d^2} + \boldsymbol{K}_{d,d})(\boldsymbol{\omega}(\tau)\otimes\boldsymbol{I}_d)$. Let $\boldsymbol{D}_1$ be the duplication matrix such that $\text{vec}[\boldsymbol{\Omega}(\tau)] = \boldsymbol{D}_1\text{vech}[\boldsymbol{\Omega}(\tau)]$, which follows that $\boldsymbol{N}_1(\tau)\text{vec}(\text{d}\boldsymbol{\omega}(\tau)) = \boldsymbol{D}_1\text{d}\boldsymbol{\sigma}(\tau)$ and

$$\boldsymbol{N}_1(\tau)\nabla_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau) = \boldsymbol{D}_1. \tag{B.5}$$

We then illustrate how to combine equation (B.5) 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 $[\boldsymbol{0}, \nabla_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]$. Let $\boldsymbol{L}_d$ be the elimination matrix defined by $\text{vech}[\boldsymbol{\omega}(\tau)] = \boldsymbol{L}_d\text{vec}[\boldsymbol{\omega}(\tau)]$. Because $\boldsymbol{\omega}(\tau)$ is lower triangular subject to short-run restrictions, $\boldsymbol{L}^\top$

is a duplication matrix such that $\text{vec}[\boldsymbol{\omega}(\tau)] = \boldsymbol{L}_d^\top \text{vech}[\boldsymbol{\omega}(\tau)]$. Write

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

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

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

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

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

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 (B.5) must be extended in the form $\boldsymbol{N}_1 \nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) = [\boldsymbol{0}, \boldsymbol{D}_1]$. Second, long-run restrictions can be expressed as $\boldsymbol{Q}\text{vec}\left[\boldsymbol{A}_\tau^{-1}(1)\boldsymbol{\omega}\right] = 0$, where $\boldsymbol{Q}$ is a $d(d-1)/2 \times d^2$ matrix of 0 and 1, and $\boldsymbol{A}_\tau(1) = \boldsymbol{I}_d - \sum_{i=1}^p \boldsymbol{A}_i(\tau)$. By $\text{d}\boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1} \cdot \text{d}\boldsymbol{A} \cdot \boldsymbol{A}^{-1}$, we have

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

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

$$
\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) = [\nabla_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\omega}(\tau), \nabla_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]
$$

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

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

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

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

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

The proof is now completed. □

Proof of Theorem 3.2.

We first provide a joint central limit theory for $\mathrm{vec}[\widehat{\boldsymbol{A}}(\tau)]$ and $\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau)$, and then prove this theorem by using Delta method.

Let $Q_T^{-1}(\tau) = [\boldsymbol{I}_d, \boldsymbol{0}_d] \left\{ \frac{1}{T}\sum_{t=1}^T (\boldsymbol{W}_t^* - \boldsymbol{W}_\tau^\top \boldsymbol{s}^\top(\tau_t)\boldsymbol{Z}_{t-1}))(\boldsymbol{X}_{1,t}^{*,\top} - \boldsymbol{Z}_{t-1}^\top \boldsymbol{s}(\tau_t)\boldsymbol{X}_\tau)K_h(\tau_t - \tau) \right\}^{-1}$. We then have

$$\begin{aligned}
\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) &= Q_T^{-1}(\tau) \times \frac{1}{T}\sum_{t=1}^T (\boldsymbol{W}_t^* - \boldsymbol{W}_\tau^\top \boldsymbol{s}^\top(\tau_t)\boldsymbol{Z}_{t-1}))(\boldsymbol{x}_t - \boldsymbol{Z}_{t-1}^\top \boldsymbol{s}(\tau_t)\boldsymbol{x})K_h(\tau_t - \tau) \\
&= Q_T^{-1}(\tau) \times \frac{1}{T}\sum_{t=1}^T (\boldsymbol{W}_t^* - \boldsymbol{W}_\tau^\top \boldsymbol{s}^\top(\tau_t)\boldsymbol{Z}_{t-1}))(\boldsymbol{X}_{1,t}\boldsymbol{\omega}_{\bullet,1}^*(\tau_t) - \boldsymbol{Z}_{t-1}^\top \boldsymbol{s}(\tau_t) \begin{bmatrix} \boldsymbol{X}_{1,1}\boldsymbol{\omega}_{\bullet,1}^*(\tau_t) \\ \vdots \\ \boldsymbol{X}_{1,T}\boldsymbol{\omega}_{\bullet,1}^*(\tau_t) \end{bmatrix})K_h(\tau_t - \tau) \\
&\quad - Q_T^{-1}(\tau) \times \frac{1}{T}\sum_{t=1}^T (\boldsymbol{W}_t^* - \boldsymbol{W}_\tau^\top \boldsymbol{s}^\top(\tau_t)\boldsymbol{Z}_{t-1}))\boldsymbol{Z}_{t-1}^\top \boldsymbol{s}(\tau_t) \begin{bmatrix} \boldsymbol{X}_{1,1}(\boldsymbol{\omega}_{\bullet,1}^*(\tau_1) - \boldsymbol{\omega}_{\bullet,1}^*(\tau_t)) \\ \vdots \\ \boldsymbol{X}_{1,T}(\boldsymbol{\omega}_{\bullet,1}^*(\tau_T) - \boldsymbol{\omega}_{\bullet,1}^*(\tau_t)) \end{bmatrix} K_h(\tau_t - \tau) \\
&\quad + Q_T^{-1}(\tau) \cdot \frac{1}{T}\sum_{t=1}^T (\boldsymbol{W}_t^* - \boldsymbol{W}_\tau^\top \boldsymbol{s}^\top(\tau_t)\boldsymbol{Z}_{t-1}))
\end{aligned}$$

$$
\times\left(\boldsymbol{Z}_{t-1}^{\top}\mathrm{vec}(\boldsymbol{A}^*(\tau_t)) - \boldsymbol{Z}_{t-1}^{\top}\boldsymbol{s}(\tau_t)\begin{bmatrix}\boldsymbol{Z}_0^{\top}\mathrm{vec}(\boldsymbol{A}^*(\tau_1))\\ \vdots \\ \boldsymbol{Z}_{T-1}^{\top}\mathrm{vec}(\boldsymbol{A}^*(\tau_T))\end{bmatrix}\right)K_h(\tau_t-\tau)
$$

$$
+Q_T^{-1}(\tau)\times\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{W}_t^* - \boldsymbol{W}_{\tau}^{\top}\boldsymbol{s}^{\top}(\tau_t)\boldsymbol{Z}_{t-1}))(\boldsymbol{\eta}_t^* - \boldsymbol{Z}_{t-1}^{\top}\boldsymbol{s}(\tau_t)\boldsymbol{\eta}^*)K_h(\tau_t-\tau)
$$

$$
:= I_{T,1} + I_{T,2} + I_{T,3} + I_{T,4},
$$

where $\boldsymbol{W}_t^* = \boldsymbol{W}_t \otimes [1, (\tau_t-\tau)/h]^{\top}$, $\boldsymbol{X}_{1,t}^* = \boldsymbol{X}_{1,t}\otimes[1,(\tau_t-\tau)/h]^{\top}$ and $\boldsymbol{\eta}^* = [\boldsymbol{\eta}_1^{*,\top}, ..., \boldsymbol{\eta}_T^{*,\top}]^{\top}$.

By Lemma B.8, we have $I_{T,2} = o_P(1/\sqrt{Th})$ and $I_{T,3} = o_P(1/\sqrt{Th})$. By using standard local linear arguments and Lemma B.8.3, we have $I_{T,1} = \boldsymbol{\omega}_{\bullet,1}^*(\tau) + \frac{1}{2}h^2\boldsymbol{\omega}_{\bullet,1}^{*,(2)}(\tau) + o_P(h^2)$. Since $\sup_{\tau\in[0,1]}\|\boldsymbol{s}(\tau)\boldsymbol{\eta}^*\| = O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right)$ by Lemma B.6.1, and by similar arguments to the proof of Lemma B.8.5, for $\tau\in(0,1)$, we have

$$
\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) - \boldsymbol{\omega}_{\bullet,1}^*(\tau) - \frac{1}{2}h^2\boldsymbol{\omega}_{\bullet,1}^{*,(2)}(\tau)
$$

$$
= (\Sigma^{*,-1}(\tau)\otimes\boldsymbol{I}_d)\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{W}_t - \boldsymbol{\Sigma}_{\boldsymbol{ZW}}^{\top}(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t)\boldsymbol{Z}_{t-1}))\boldsymbol{\eta}_t^*K_h(\tau_t-\tau) + o_P(1/\sqrt{Th}).
$$

Hence, we have

$$
\sqrt{Th}\begin{bmatrix}\mathrm{vec}\left(\widehat{\boldsymbol{A}}(\tau) - \boldsymbol{A}(\tau) - \frac{1}{2}h^2\tilde{c}_2\boldsymbol{A}^{(2)}(\tau)\right)\\ \widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) - \boldsymbol{\omega}_{\bullet,1}^*(\tau) - \frac{1}{2}h^2\boldsymbol{\omega}_{\bullet,1}^{*,(2)}(\tau)\end{bmatrix}
$$

$$
= \begin{bmatrix}(\boldsymbol{\Sigma}^{-1}(\tau)\otimes\boldsymbol{I}_d)\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{\eta}_tK\left(\frac{\tau_t-\tau}{h}\right)\right)\\ (\Sigma^{*,-1}(\tau)\otimes\boldsymbol{I}_d)\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}(\boldsymbol{W}_t - \boldsymbol{\Sigma}_{\boldsymbol{ZW}}^{\top}(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t)\boldsymbol{Z}_{t-1})\boldsymbol{\eta}_t^*K\left(\frac{\tau_t-\tau}{h}\right)\right)\end{bmatrix} + o_P(1)
$$

$$
:= I_{T,5} + o_P(1).
$$

By simple calculation, we have

$$
Var(I_{T,5}) \to \begin{bmatrix}\boldsymbol{V}_{1,1}(\tau) & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{V}_{2,2}^*(\tau)\end{bmatrix},
$$

where $\boldsymbol{V}_{2,2}^*(\tau) = (\Sigma^{*,-2}(\tau)\Delta^*(\tau))\otimes\boldsymbol{\Omega}^*(\tau)$, $\Delta^*(\tau) = \Sigma_{\pi}(\tau) - \boldsymbol{\Sigma}_{\boldsymbol{z}\pi}^{\top}(\tau)\boldsymbol{\Sigma}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{z}\pi}(\tau)$ and $\boldsymbol{\Omega}^{IV}(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E(\boldsymbol{\eta}_t^*\boldsymbol{\eta}_t^{*,\top})K_h(\tau_t-\tau)$. In addition, similar to the proof of Theorem 2.1, we can prove the asymptotic normality of $I_{T,5}$ by using martingale central limit theorem.

Next, similar to the proof of Theorem 3.1, we have

$$
\nabla_{\boldsymbol{\alpha}(\tau)}\boldsymbol{B}_{,\cdot 1}(\tau) = \frac{\partial\mathrm{vec}\left[\boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}_{\bullet,1}^*(\tau)\right]}{\partial\boldsymbol{\alpha}^{\top}(\tau)} = (\boldsymbol{\omega}_{\bullet,1}^{*,\top}(\tau)\otimes\boldsymbol{I}_d)\frac{\partial\mathrm{vec}\left[\boldsymbol{\Psi}_j(\tau)\right]}{\partial\boldsymbol{\alpha}^{\top}(\tau)}
$$

$$
= (\boldsymbol{\omega}^{*,\top}_{\bullet,1}(\tau) \otimes \boldsymbol{I}_d) \frac{\partial \mathrm{vec}\left[\boldsymbol{J}\boldsymbol{\Phi}^j(\tau)\boldsymbol{J}^\top\right]}{\partial \boldsymbol{\alpha}^\top(\tau)}
$$

$$
= (\boldsymbol{\omega}^{*,\top}_{\bullet,1}(\tau) \otimes \boldsymbol{I}_d) \left(\sum_{m=0}^{j-1} \boldsymbol{J}(\boldsymbol{\Phi}^\top(\tau))^{j-1-m} \otimes \boldsymbol{\Psi}_m(\tau)\right)
$$

and $\nabla_{\boldsymbol{\omega}^{*,\top}_{\bullet,1}(\tau)} \boldsymbol{B}_{,\cdot 1}(\tau) = \boldsymbol{\Psi}_j(\tau)$.

Similar to the proof of Theorem 3.1, the result is obtained by the Delta method. □

B.3 Proofs of the Preliminary Lemmas

Proof of Lemma B.5.

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

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

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

$$
\sum_{t=1}^T \left\|\boldsymbol{x}_t\boldsymbol{x}_t^\top - \widetilde{\boldsymbol{x}}_t\widetilde{\boldsymbol{x}}_t^\top\right\| \leq \sum_{t=1}^T (\|\boldsymbol{x}_t\| + \|\widetilde{\boldsymbol{x}}_t\|)\|\boldsymbol{x}_t - \widetilde{\boldsymbol{x}}_t\|
$$

$$
\leq \left(\sum_{t=1}^T (\|\boldsymbol{x}_t\| + \|\widetilde{\boldsymbol{x}}_t\|)^2\right)^{1/2} \left(\sum_{t=1}^T \|\boldsymbol{x}_t - \widetilde{\boldsymbol{x}}_t\|^2\right)^{1/2} = O_P(\sqrt{T}) \cdot O_P(1/\sqrt{T}) = O_P(1).
$$

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

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

and

$$
\sum_{t=1}^{T-1}\sum_{j=0}^\infty j\,\|\boldsymbol{B}_j(\tau_{t+1}) - \boldsymbol{B}_j(\tau_t)\|
$$

$$
\leq \sup_{\tau\in[0,1]} \sum_{j=1}^\infty j\|\boldsymbol{\Psi}_j(\tau)\| \sum_{t=1}^T \|\boldsymbol{\omega}(\tau_{t+1}) - \boldsymbol{\omega}(\tau_t)\| + M \sum_{t=1}^{T-1}\sum_{j=0}^\infty j\|\boldsymbol{\Psi}_j(\tau_{t+1}) - \boldsymbol{\Psi}_j(\tau_t)\|
$$

$$
\leq M \sum_{j=0}^\infty (j^2\rho_A^{j-1} + j\rho_A^j) < \infty.
$$

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). □

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,

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

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

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

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

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

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

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

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

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

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

which follows from

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

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

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

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

(2). Note that

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

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 $\boldsymbol{S}_T^{-1}(\tau_t)$ and hence,

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

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

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

For $J_{T,321}$,

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

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

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

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

$$\begin{aligned}
J_{T,322} &= \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\boldsymbol{\mu}_{t-1}^\top \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j,t+i-1} \boldsymbol{\eta}_{t+i-1-j}\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&\quad + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \boldsymbol{\eta}_{t-1-j}^\top \boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\mu}_{t+i-1}\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&\quad + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \left(\boldsymbol{\eta}_{t-1-j}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top - E\left(\boldsymbol{\eta}_{t-1-j}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top\right)\right)\right. \\
&\quad \left. \cdot \operatorname{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{j+i,t+i-1}\right)\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&\quad + \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \sum_{m=0, \neq j+i}^{\infty} \left(\boldsymbol{\eta}_{t+i-1-m}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top\right) \operatorname{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{m,t+i-1}\right)\right) \\
&\quad \cdot \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&:= J_{T,3221} + J_{T,3222} + J_{T,3223} + J_{T,3224}.
\end{aligned}$$

For $J_{T,3221}$,

$$E\left\|J_{T,3221}\right\|^2$$

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

$$
= O\left(\frac{1}{Th}\right).
$$

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

$$
J_{T,3224} = \frac{1}{(Th)^{3/2}}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{\eta}_t\left(\sum_{j=0}^{\infty}\left(\boldsymbol{\eta}_t^\top\otimes\boldsymbol{\eta}_{t-1-j}^\top\right)\mathrm{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top\boldsymbol{\Psi}_{i-1,t+i-1}\right)\right)\boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right)K\left(\frac{\tau_t-\tau}{h}\right)
$$

$$
+\frac{1}{(Th)^{3/2}}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{\eta}_t\left(\sum_{j=0}^{\infty}\sum_{m=0,\neq j+i,\neq i-1}^{\infty}\left(\boldsymbol{\eta}_{t+i-1-m}^\top\otimes\boldsymbol{\eta}_{t-1-j}^\top\right)\right.
$$

$$
\left.\cdot\,\mathrm{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top\boldsymbol{\Psi}_{m,t+i-1}\right)\right)\boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right)K\left(\frac{\tau_t-\tau}{h}\right) := J_{T,32241}+J_{T,32242}.
$$

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

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

$$
E\left\|J_{T,32241}\right\|^2 = \frac{1}{(Th)^3}\sum_{i_1=1}^{T-1}\sum_{i_2=1}^{T-1}\sum_{t_1=1}^{T-i_1}E\left\|\boldsymbol{\eta}_{t_1+i_1}^\top\boldsymbol{\eta}_{t_1+i_1}\right\|E\left\|\boldsymbol{w}_{t_1+i_1-i_2,i_2}^\top\boldsymbol{w}_{t_1,i_1}\right\|
$$

$$
\cdot K\left(\frac{i_1}{Th}\right)K\left(\frac{i_2}{Th}\right)K\left(\frac{\tau_{t_1}-\tau}{h}\right)K\left(\frac{\tau_{t_1+i_1-i_2}-\tau}{h}\right)
$$

$$
\le \frac{M}{(Th)^3}\sum_{t_1=1}^{T}\left(\max_t\sum_{i=1}^{T-1}\left\|\boldsymbol{\Psi}_{i,t}\right\|\right)^2\left(\max_t\sum_{j=0}^{\infty}\left\|\boldsymbol{\Psi}_{j,t}\right\|\right)^2 K\left(\frac{\tau_{t_1}-\tau}{h}\right) = O\left((Th)^{-2}\right).
$$

Hence, $J_{T,32} = O_P\left((Th)^{-1/2}\right)$. Similar to $J_{T,32}$, $J_{T,33} = O_P\left((Th)^{-1/2}\right)$. 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

$$
\widehat{\boldsymbol{\Omega}}(\tau) = \frac{1}{Th}\sum_{t=1}^{T}\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top K\left(\frac{\tau_t-\tau}{h}\right)
$$

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

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

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

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

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

Hence, by the first result of Theorem 2.1

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

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

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

The proof is now completed. □

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

Proof of Lemma B.7.

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

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

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,

$$\begin{aligned}I_{T,1} &\leq \frac{1}{T}\sum_{t=1}^T \|\boldsymbol{z}_{\mathsf{p},t-1}\|^2 \cdot \left\|\widehat{\mathbf{\Lambda}}_{\mathsf{p}}(\tau_t) - \mathbf{\Lambda}_{\mathsf{p}}(\tau_t)\right\|^2 \leq \sup_{0\leq\tau\leq1}\left\|\widehat{\mathbf{\Lambda}}_{\mathsf{p}}(\tau) - \mathbf{\Lambda}_{\mathsf{p}}(\tau)\right\|^2 \cdot \frac{1}{T}\sum_{t=1}^T \|\boldsymbol{z}_{\mathsf{p},t-1}\|^2 \\ &= O_P\left((h^2 + (\log T/(Th))^{1/2})^2\right).\end{aligned}$$

For $I_{T,2}$,

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_t\left(\boldsymbol{\eta}_t - \widehat{\boldsymbol{\eta}}_{\mathsf{p},t}\right)^\top = \frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\left(\widehat{\mathbf{\Lambda}}_{\mathsf{p}}(\tau_t) - \mathbf{\Lambda}_{\mathsf{p}}(\tau_t)\right)^\top \\ &= \frac{1}{2}h^2 \cdot \frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_t[\boldsymbol{z}_{\mathsf{p},t-1}^\top, \mathbf{0}_{(d^2\mathsf{p}+d)\times1}^\top]\boldsymbol{S}_T^{-1}(\tau_t)\begin{pmatrix}\boldsymbol{S}_{T,2}(\tau_t) \\ \boldsymbol{S}_{T,3}(\tau_t)\end{pmatrix}\boldsymbol{A}_{\mathsf{p}}^{(2),\top}(\tau_t) \\ &+\frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_t[\boldsymbol{z}_{\mathsf{p},t-1}^\top, \mathbf{0}_{(d^2\mathsf{p}+d)\times1}^\top]\boldsymbol{S}_T^{-1}(\tau_t)\left(\frac{1}{Th}\sum_{s=1}^T\left(\boldsymbol{z}_{\mathsf{p},s-1}^*\boldsymbol{z}_{\mathsf{p},s-1}^\top \otimes \boldsymbol{I}_d\right)\boldsymbol{M}_{\mathsf{p}}^\top(\tau_s)K\left(\frac{\tau_s - \tau_t}{h}\right)\right) \\ &+\frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_t[\boldsymbol{z}_{\mathsf{p},t-1}^\top, \mathbf{0}_{(d^2\mathsf{p}+d)\times1}^\top]\boldsymbol{S}_T^{-1}(\tau_t)\left(\frac{1}{Th}\sum_{s=1}^T \boldsymbol{Z}_{\mathsf{p},s-1}^*\boldsymbol{\eta}_s^\top K\left(\frac{\tau_s - \tau_t}{h}\right)\right) := I_{T,3} + I_{T,4} + I_{T,5}.\end{aligned}$$

By the uniform convergence results stated in Lemmas B.5.2–3, we replace the weighed sample covariance with its limit plus the rate $O_P\left((\log T/(Th))^{1/2}\right)$, 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 $\mathbf{\Sigma}(\tau) = \mathrm{plim}_{T\to\infty}\boldsymbol{S}_{T,0}(\tau)$, we have

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

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

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

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

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

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

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

Proof of Lemma B.8.

(1). By Lemma B.5, uniformly over $\tau\in[h,1-h]$, we have

$$\begin{aligned}\boldsymbol{s}(\tau)\boldsymbol{X}_\tau&=[\boldsymbol{I}_{d^2p+d},\boldsymbol{0}_{d^2p+d}]\begin{pmatrix}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}^{\top}K_h(\tau_t-\tau) & \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}^{\top}\left(\frac{\tau_t-\tau}{h}\right)K_h(\tau_t-\tau)\\ \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}^{\top}\left(\frac{\tau_t-\tau}{h}\right)K_h(\tau_t-\tau) & \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}^{\top}\left(\frac{\tau_t-\tau}{h}\right)^2K_h(\tau_t-\tau)\end{pmatrix}^{-1}\\&\quad\times\begin{pmatrix}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{X}_{1,t}K_h(\tau_t-\tau) & \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{X}_{1,t}\left(\frac{\tau_t-\tau}{h}\right)K_h(\tau_t-\tau)\\ \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{X}_{1,t}\left(\frac{\tau_t-\tau}{h}\right)K_h(\tau_t-\tau) & \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{X}_{1,t}\left(\frac{\tau_t-\tau}{h}\right)^2K_h(\tau_t-\tau)\end{pmatrix}\\&=\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{X}_1}(\tau)\otimes[1,0]+O_P\left(h^2+\left(\frac{\log T}{Th}\right)^{1/2}\right).\end{aligned}$$

(2). Similar to the proof of part (1), by Assumption 4, part (2) is easily obtained by Lemma B.5.

(3). By parts (1)–(2) and Lemma B.5, the first $d\times d$ matrix of $\frac{1}{T}\boldsymbol{W}_\tau^{\top}(\boldsymbol{I}_{dT}-\boldsymbol{S})^{\top}\boldsymbol{K}_\tau(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{X}_\tau$ is

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{W}_t-\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{W}}^{\top}(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t)\boldsymbol{Z}_{t-1})(\boldsymbol{X}_{1,t}-\boldsymbol{Z}_{t-1}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{X}}(\tau_t))K_h(\tau_t-\tau)+O_P\left(h^2+\left(\frac{\log T}{Th}\right)^{1/2}\right)\\&=(\Sigma_{\pi x_1}(\tau)-\boldsymbol{\Sigma}_{\boldsymbol{z}\pi}^{\top}(\tau)\boldsymbol{\Sigma}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{z}x_1}(\tau))\otimes\boldsymbol{I}_d+O_P\left(h^2+\left(\frac{\log T}{Th}\right)^{1/2}\right)\end{aligned}$$

uniformly over $\tau\in[h,1-h]$. The proofs of the other components in $\frac{1}{T}\boldsymbol{W}_\tau^{\top}(\boldsymbol{I}_{dT}-\boldsymbol{S})^{\top}\boldsymbol{K}_\tau(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{X}_\tau$ are similar, so omitted here.

(4). Let $\rho_T = h^2 + \left(\frac{\log T}{Th}\right)^{1/2}$. By Lemma B.5, we have

$$[\boldsymbol{Z}_{t-1}^\top, \mathbf{0}_{d^2p+d}](\boldsymbol{Z}_{\tau_t}^\top \boldsymbol{K}_{\tau_t} \boldsymbol{Z}_{\tau_t})^{-1} \boldsymbol{Z}_{\tau_t}^\top \boldsymbol{K}_{\tau_t} \begin{bmatrix} \boldsymbol{Z}_0^\top \mathrm{vec}(\boldsymbol{A}^*(\tau_1)) \\ \vdots \\ \boldsymbol{Z}_{T-1}^\top \mathrm{vec}(\boldsymbol{A}^*(\tau_T)) \end{bmatrix} = \boldsymbol{Z}_{t-1}^\top \mathrm{vec}(\boldsymbol{A}^*(\tau_t))(1 + O_P(\rho_T))$$

uniform over $t \in [1, 2..., T]$. Hence, by Lemma B.5, we have

$$\begin{aligned} &\boldsymbol{W}_\tau^\top (\boldsymbol{I}_{dT} - \boldsymbol{S})^\top \boldsymbol{K}_\tau (\boldsymbol{I}_{dT} - \boldsymbol{S}) \begin{bmatrix} \boldsymbol{Z}_0^\top \mathrm{vec}(\boldsymbol{A}^*(\tau_1)) \\ \vdots \\ \boldsymbol{Z}_{T-1}^\top \mathrm{vec}(\boldsymbol{A}^*(\tau_T)) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{t=1}^T \left(\boldsymbol{W}_t - \boldsymbol{\Sigma}_{\boldsymbol{ZW}}^\top(\tau)\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau)\boldsymbol{Z}_{t-1}(1 + O_P(\rho_T))\right) \boldsymbol{Z}_{t-1}^\top \mathrm{vec}(\boldsymbol{A}^*(\tau_t)) O_P(\rho_T) K_h(\tau_t - \tau) \\ \sum_{t=1}^T \left(\boldsymbol{W}_t \left(\frac{\tau_t - \tau}{h}\right) + O_P(\rho_T)\right) \boldsymbol{Z}_{t-1}^\top \mathrm{vec}(\boldsymbol{A}^*(\tau_t)) O_P(\rho_T) K_h(\tau_t - \tau) \end{bmatrix} \\ &= O_P(Th\rho_T^2) = o_P(\sqrt{Th}). \end{aligned}$$

(5). Similar to the proof of part (1), we have

$$\boldsymbol{s}(\tau_t) \begin{bmatrix} \boldsymbol{X}_{1,1}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_1) - \boldsymbol{\omega}_{\cdot,1}^*(\tau_t)) \\ \vdots \\ \boldsymbol{X}_{1,T}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_T) - \boldsymbol{\omega}_{\cdot,1}^*(\tau_t)) \end{bmatrix} = \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t) \boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{X}_1}(\tau_t) O_P(\rho_T)$$

uniformly over $t = 1, 2, ..., T$. Hence, we have

$$\begin{aligned} &\frac{1}{\sqrt{Th}} \sum_{t=1}^T [\boldsymbol{W}_t - \boldsymbol{\Sigma}_{\boldsymbol{ZW}}^\top(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t)\boldsymbol{Z}_{t-1}(1 + O_P(\rho_T))] \boldsymbol{Z}_{t-1}^\top \boldsymbol{s}(\tau_t) \begin{bmatrix} \boldsymbol{X}_{1,1}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_1) - \boldsymbol{\omega}_{\cdot,1}^*(\tau_t)) \\ \vdots \\ \boldsymbol{X}_{1,T}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_T) - \boldsymbol{\omega}_{\cdot,1}^*(\tau_t)) \end{bmatrix} K_h(\tau_t - \tau) \\ &= \frac{1}{\sqrt{Th}} \sum_{t=1}^T [\boldsymbol{W}_t - \boldsymbol{\Sigma}_{\boldsymbol{ZW}}^\top(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t)\boldsymbol{Z}_{t-1}(1 + O_P(\rho_T))] \boldsymbol{Z}_{t-1}^\top \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\tau_t) \boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{X}_1}(\tau_t) O_P(\rho_T) K_h(\tau_t - \tau) \\ &= O_P(\sqrt{Th}\rho_T^2) = o_P(1). \end{aligned}$$

The proof is now completed. □

Proof of Lemma B.9.

(1). Write

$$\sigma^2 = \frac{1}{T^2 h} \sum_{t=2}^T \sum_{s=1}^{t-1} \left[\int_{-1}^1 K(u) K\left(u + \frac{t-s}{Th}\right) \mathrm{d}u\right]^2$$

$$
\begin{aligned}
&= \frac{1}{T^2 h}\sum_{t=2}^{T}\sum_{j=1}^{t-1}\left[\int_{-1}^{1} K(u)\, K\left(u+\frac{j}{Th}\right) \mathrm{d}u\right]^2 \\
&= \frac{1}{Th}\sum_{j=1}^{T-1}(1-j/T)\left[\int_{-1}^{1} K(u)\, K\left(u+\frac{j}{Th}\right) \mathrm{d}u\right]^2 \\
&= \int_{0}^{\infty}(1-vh)\left[\int_{-1}^{1} K(u)\, K(u+v)\, \mathrm{d}u\right]^2 \mathrm{d}v + O(1/(Th)) \\
&\to \int_{0}^{2}\left[\int_{-1}^{1-v} K(u)\, K(u+v)\, \mathrm{d}u\right]^2 \mathrm{d}v.
\end{aligned}
$$

(2). Write

$$
\begin{aligned}
\max_t |a_t| &= \max_t \left|\sum_{i=1}^{t-1}\frac{1}{T^2 h}\left[\int_{-1}^{1} K(u)\, K\left(u+\frac{i}{Th}\right) \mathrm{d}u\right]^2\right| \\
&\leq \frac{1}{T^2 h}\sum_{i=1}^{T}\left[\int_{-1}^{1} K(u)\, K\left(u+\frac{i}{Th}\right) \mathrm{d}u\right]^2 \\
&= \frac{1}{T}\int_{0}^{\infty}\left[\int_{-1}^{1} K(u)\, K(u+v)\, \mathrm{d}u\right]^2 \mathrm{d}v(1+o(1)) \\
&= O(1/T).
\end{aligned}
$$

(3). Write

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

(4). Write

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

(5). By Cauchy–Schwarz inequality,

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

□

Proof of Lemma B.10.

Write $\boldsymbol{\xi}_t = [\xi_{t,1}, ..., \xi_{t,d}]^\top$. We first prove

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

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

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

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

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

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

Proof of Lemma B.11.

Without loss of generality, let $E(\boldsymbol{w}_t^*) = \mathbf{0}$. For any integer $I \geq 1$ introduce the truncated process $\boldsymbol{h}_{t-1,I}^* = E\left(\boldsymbol{h}_{t-1}^* | \mathcal{F}_{t-I}\right)$. Then $\boldsymbol{h}_{t-1,I}^* = 0$ if $t \leq I$ and $\boldsymbol{h}_{t-1,I}^* = \sum_{s=1}^{t-I} w_{s,t}\boldsymbol{y}_s$ for $1 \leq I < t$. For $2 \leq t \leq T$, by Lemma B.10,

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

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

$$T(I) = \sum_{t=2}^{T} \operatorname{tr}\left[E(\boldsymbol{w}_t^* | \mathcal{F}_{t-I}) \boldsymbol{h}_{t-1,I}^* \boldsymbol{h}_{t-1,I}^{*,\top}\right].$$

By Cauchy–Schwarz inequality, Lemma B.9 (iii), if $I/(Th) \to 0$, we have

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

since

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

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

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

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

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

By Lemma B.9 (ii), Lemma B.10 and Cauchy–Schwarz inequality, since $|||\mathcal{P}_{t-j}\boldsymbol{w}_t^*|||_{\delta/2} \le 2|||\boldsymbol{w}_t^*|||_{\delta/2} < \infty$,

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

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

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

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

Proof of Lemma B.12.

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

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

Consider $I_{T,1}$. Write

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

$$
+\frac{1}{T}\sum_{s=1}^{T-1}\text{tr}\left[\boldsymbol{\Omega}(\tau_s)\left(\boldsymbol{Z}_{s-1}^\top\boldsymbol{Z}_{s-1}-E\left(\boldsymbol{Z}_{s-1}^\top\boldsymbol{Z}_{s-1}\right)\right)\right]\left(T\sum_{t=s+1}^{T}w_{s,t}^2\right)
$$

$$
= I_{T,11}+I_{T,12}.
$$

Since $\text{tr}\left[\left(\boldsymbol{\eta}_s\boldsymbol{\eta}_s^\top-\boldsymbol{\Omega}(\tau_s)\right)\boldsymbol{Z}_{s-1}^\top\boldsymbol{Z}_{s-1}\right]$, $s=1,2,...$ are a martingale difference sequence and $T\sum_{t=1}^{T}w_{s,t}^2=O(1)$ by Lemma B.9.2, 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.10, Cauchy–Schwarz inequality and Lemma B.9.5,

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

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

Proof of Lemma B.13. Note that $\boldsymbol{y}_t^\top\boldsymbol{H}_t\boldsymbol{h}_{t-1}^*$, $t\in\mathbb{Z}$, 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

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

By Cauchy–Schwarz inequality, Lemma B.10 and Lemma B.9 (2), the Lindeberg condition is satisfied since

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

Apply Lemmas B.11 and B.12 with $\boldsymbol{w}_t^*=E\left(\boldsymbol{H}_t^\top\boldsymbol{y}_t\boldsymbol{y}_t^\top\boldsymbol{H}_t|\mathcal{F}_{t-1}\right)$, then we have the convergence of conditional variance

$$
\sum_{t=2}^{T}\text{tr}\left[E\left(\boldsymbol{H}_t^\top\boldsymbol{y}_t\boldsymbol{y}_t^\top\boldsymbol{H}_t|\mathcal{F}_{t-1}\right)\boldsymbol{h}_{t-1}^*\boldsymbol{h}_{t-1}^{*,\top}\right]
$$

$$\to_P \sum_{t=2}^{T} \mathrm{tr}\left[E\left(\boldsymbol{H}_t^\top \boldsymbol{y}_t \boldsymbol{y}_t^\top \boldsymbol{H}_t\right) E\left(\boldsymbol{h}_{t-1}^* \boldsymbol{h}_{t-1}^{*,\top}\right)\right].$$

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

□

Proof of Lemma B.14.

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 $Th^6 \to 0$, we have

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

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

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

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

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

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

by Lemma B.1. The proof is now completed. □