On Multivariate Time–Varying Dynamic Models

Yayi Yan

Master of Economics, Nankai University, 2019

Bachelor of Economics, Huazhong University of Science and Technology, 2017

A thesis submitted for the degree of

Doctor of Philosophy

at Monash University in 2022

Department of Econometrics and Business Statistics

Faculty of Business and Economics

Monash University

2022

Copyright notice

Abstract

This dissertation consists of three chapters that contribute to different multivariate time series models with local stationarity; that is, the underlying data generating mechanism of the dynamic process is changing smoothly over time. The first chapter briefly reviews the literature. The second chapter considers a new class of time-varying vector moving average infinity (VMA(∞)) processes. Accordingly, some new asymptotic properties are established, including the law of large numbers, uniform convergence, the central limit theory, bootstrap consistency, and long-run covariance matrix estimation for the class of time-varying VMA(∞) processes. The third chapter introduces a new class of time-varying vector autoregression (VAR) models in which the VAR coefficients and covariance matrix of the error innovations are allowed to change smoothly over time. Accordingly, a set of asymptotic properties are established, 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. The fourth chapter considers a wide class of time-varying multivariate causal processes that nests many classic and new examples as special cases. The existence of a weakly dependent stationary approximation for the model is considered, which is the foundation to initiate the theoretical development. Further, I consider the quasi-maximum likelihood estimation (QMLE) approach, and provide both point-wise and simultaneous inferences on the functional time-varying coefficients. Numerical studies are conducted to illustrate the usefulness of the proposed models and methods.

Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma at any university or equivalent institution and, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Yayi Yan

28 June 2022

Acknowledgements

Firstly, I would like to thank my supervisors, who have provided me with wisdom throughout this journey. In particular, I would like to thank Jiti Gao and Bin Peng, for suggesting this research topic, as well as for their efforts and patience.

To my postgraduate peers, the last few years would not have been possible without your friendship and support. In particular, to my office mates, the many conversations and social outings we enjoyed together helped keep me sail through the years.

I would also like to thank my parents, who have wholeheartedly and unconditionally supported me in all my endeavours throughout my life. I truly believe that without their unwavering support, I would not be here today.

Thank you all.

Contents

Chapter 1

Introduction

Studying the joint behavior of multivariate time series is one of the most fundamental tasks in the fields of statistics and econometrics. In this direction, a variety of multivariate time series models have been proposed and are widely adopted in different disciplines. For example, see Fan et al. [2008], Bardet and Wintenberger [2009], and McElroy and Roy [2021] for extensive discussions on different vector autoregressions (VAR) and multivariate generalised autoregressive conditional heteroscedasticity (GARCH) models. The majority of these multivariate models are constructed assuming that the underlying time series is stationary, which in a sense is very restrictive given different macro shocks, time trends, and changes in volatility in practice. For instance, economic and financial data always include different macro shocks. As a result, the behaviour can be rather volatile; for instance, the climate data may contain a certain time trend that recently has attracted significant attention due to greenhouse emission. Anyway, certain nonstationarity always occurs.

To account for nonstationarity, locally stationary processes have received considerable attention since the seminal work of Dahlhaus [1996b], Dette et al. [2011], Zhang and Wu [2012], Truquet [2017], and Dahlhaus et al. [2019], among others. The type of processes which can be described with this infill asymptotics are processes which locally at each time point are close to a stationary process but whose characteristics (covariances, parameters, etc.) are gradually changing in an unspecific way as time evolves. The simplest example for such a process may be an AR(p)-process whose parameters are varying in time. In contrast to the unit root process, the locally stationary process nicely balances stationarity and nonstationarity by allowing for the simultaneous presence of both types of behaviours in one time series process. It is worth pointing out that although nonparametric estimation for deterministic time–varying models has received much attention initially on time series regression models (Robinson, 1989, Cai, 2007, Zhang and Wu, 2012, Phillips et al., 2017, Su and Wang, 2017, Li et al., 2020) over the past three decades. and then on univariate autoregressive regression models (Dahlhaus and Rao, 2006, Richter and Dahlhaus, 2019, Sun et al., 2021a, Yousuf and Ng, 2020) in recent years, very few study about multivariate autoregressive models with deterministic time–varying coefficients has been done, which somewhat limits the popularity of locally stationary processes.

To address the aforementioned issue, the literature is extended by analysing multivariate time series models with local stationarity in this dissertation. I start with investigating moving average infinity (MA(∞)) processes, which play an important role in modelling time series data. While a strand of literature on time series analysis emphasises the importance of modelling smooth changes over time and is therefore shifting its focus from parametric models to nonparametric ones, MA(∞) processes with constant parameters are often part of the fundamental data generating mechanism. Considering this line of research, an intuitive question is how to allow the underlying data generating mechanism to evolve over time. To better capture the dynamics, the second chapter of my thesis considers a new class of time-varying vector moving average infinity (VMA(∞)) processes. Accordingly, I establish some new asymptotic properties, including the law of large numbers, uniform convergence, the central limit theory, bootstrap consistency, and long-run covariance matrix estimation for the class of time-varying VMA(∞) processes. To illustrate the usefulness of the proposed method, the empirical relevance and usefulness of the newly proposed model and estimation theory are demonstrated through extensive simulated and real data studies.

The third chapter contributes to the literature on VAR models, which are widely used in practical studies, such as forecasting, modelling policy transmission mechanisms, and measuring the connection between economic agents. This chapter introduces a new class of time-varying VAR models in which the coefficient and covariance matrix of the error innovations are allowed to change smoothly over time. Accordingly, I 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, the empirical relevance and usefulness of the proposed methods are demonstrated through an application on U.S. government spending multipliers.

The fourth chapter introduces a wide class of time-varying multivariate causal processes which nests many classic and new examples as special cases. I first prove the existence of a weakly dependent stationary approximation for the model considered which is the foundation to initiate the theoretical development. Afterwards, I consider the QMLE estimation approach, and provide both point-wise and simultaneous inferences on the coefficient functions. In addition, I demonstrate the theoretical findings through both simulated and real data examples. In particular, I show the empirical relevance of this study using an application to evaluate the conditional correlations between the stock markets of China and U.S. The empirical results reveal that the interdependence between the two stock markets is increasing over time.

The fifth chapter concludes. Some preliminary lemmas and the proofs of all theoretical results are reported in the appendices of this thesis.

Before proceeding, it is essential to introduce some notations: $|\cdot|$ denotes the Euclidean norm of a vector or the spectral norm of a matrix; $\|\cdot\|_q = \{E|\cdot|^q\}^{1/q}$; $\otimes$ denotes the Kronecker product; $\mathbf{I}_a$ refers to an $a \times a$ identity matrix; $\mathbf{0}_{a\times b}$ refers to an $a \times b$ matrix of zeros, and $\mathbf{0}_a$ is short for $a = b$; for a function $g(w)$, let $g^{(j)}(w)$ be the j^{th} derivative of $g(w)$, where $j \geq 0$ and $g^{(0)}(w) \equiv g(w)$; $K_h(\cdot) = K(\cdot/h)/h$, where $K(\cdot)$ and h stand for a nonparametric kernel function and a bandwidth respectively; let $\tilde{c}_k = \int_{-1}^{1} u^k K(u)du$ and $\tilde{v}_k = \int_{-1}^{1} u^k K^2(u)du$ for integer $k \geq 0$; $\text{vec}(\cdot)$ stacks the elements of an $m \times n$ matrix as an $mn \times 1$ vector; $\text{tr}\,(\mathbf{A})$ denotes the trace of $\mathbf{A}$; let $\mathbf{A}_{i,\cdot}$ and $\mathbf{A}_{\cdot,j}$ denote the i^{th} row and the j^{th} column of matrix $\mathbf{A}$ respectively; finally, let $\to_P$ and $\to_D$ denote convergence in probability and convergence in distribution, respectively.

Chapter 2

Time-Varying Vector MA(∞) Processes

2.1 Introduction

Moving average infinity (MA(∞)) processes are possibly one of the most fundamental data generating mechanisms when studying time series (Beveridge and Nelson, 1981, Phillips and Solo, 1992, Hamilton, 1994, Lütkepohl, 2005). For example, in the field of macroeconomics, MA(∞) representations of multivariate time series with time-invariant coefficients play a central role when estimating impulse response functions, which trace the transmission mechanism of economic shocks and are useful for policy analyses. Such a routine exercise has also been widely adopted in fields such as signal processing and climatology (e.g., Bühlmann, 1998, Friedrich et al., 2020, Paul, 2020, Plagborg-Møller and Wolf, 2021). As pointed out by Hansen [2001], it seems that dynamic models with time-invariant coefficients may be unnecessarily restrictive to accommodate smooth changes over a period rather than in a static manner. To model such time-varying behaviours, an important strand of the relevant literature assumes that the coefficients of dynamic models evolve randomly (e.g., Primiceri, 2005, Petrova, 2019), and estimation procedures heavily depend on Bayesian computational algorithms.

In addition, nonparametric methods have also been proposed to estimate unknown time-varying parameters involved in some autoregressive models. It is worth noting the terminology "local stationarity", which dates back to at least the seminal work by Dahlhaus [1996a]. Since then, significant work has been undertaken to study univariate autoregressive models (e.g., Dahlhaus and Rao, 2006, Dahlhaus, 2012, Zhang and Wu, 2012, Richter and Dahlhaus, 2019). To the best of my knowledge, it has had very limited success to generalize the local stationarity approach to multivariate settings. Exceptions are the cases where different locally stationary univariate time series may be approximated by their stationary versions on the same segments, which however are not very realistic in practice. In fact, different univariate time series components of general multivariate time series may have quite different behaviours, and therefore different locally stationary univariate time series may not be approximated by their stationary versions on the same segments. To give a concrete example, see the three univariate time series plotted in Figure 2.1 of Section 2.5.3.

In order to address estimation and inferential issues for general multivariate dynamic models, I show the versatility of an alternative approach that is designed for a wide class of time-varying vector moving average (VMA)(∞) processes associated with nonparametrically unknown time-varying coefficients. In particular, I develop an explicit decomposition for partial sums of time-varying VMA(∞) processes into the long-run and transitory elements, which is known as the Beveridge-Nelson (BN) decomposition (Beveridge and Nelson, 1981, Phillips and Solo, 1992). The long-run component of the decomposition yields a martingale approximation, which ensures the feasibility of achieving a variety of asymptotic properties for multivariate dynamic models. The proposed time-varying BN decomposition then facilitates the establishment of a number of new asymptotic properties for the proposed estimators of the unknown trend functions and coefficient matrices of the general class of VMA(∞) models under some very mild assumptions. I also show that the VMA(∞) process naturally covers a class of time-varying vector autoregressive models with exogenous variables (VARX), and further establish several asymptotic properties of non- and semi-parametric estimators in time-varying VARX models. In the empirical study, I apply the newly proposed framework to study the long-run level of inflation and the natural rate of unemployment. I find that (i) the long-run level of inflation is more anchored now and is close to the Federal Reserve's target of two percent after the beginning of the Great Moderation period, and (ii) the natural rate of unemployment is less persistent and increases rapidly during "Second Oil Crisis" and "Global Financial Crisis".

In summary, my contributions in this chapter are as follows. First, I propose a new class of time-varying VMA(∞) models, and then develop a time-varying counterpart of the conventional BN decomposition before it is able to establish a variety of asymptotic properties for the estimation of the unknown trends and coefficients of the general class of time-varying VMA(∞) models, such as the law of large numbers, the uniform convergence and the central limit theory, under some very mild assumptions. Second, I propose a dependent wild bootstrap (DWB) procedure and a heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimation method to ensure that the proposed estimation theory is valid for inferential purposes and empirical implementations, and the finite-sample evaluation results show that the proposed estimation and inferential methods work well numerically. Third, after employing the time-varying BN decomposition technology, I am able to consistently estimate a class of time-varying VARX models using non- and semi-parametric methods.

This chapter is organised as follows. Section 2.2 introduces a class of time-varying VMA(∞) processes, develops a time-varying counterpart of the conventional BN decomposition, and establishes a set of asymptotic properties. Section 2.3 applies the results of Section 2.2 to establish an inferential theory for smooth deterministic trends of the time-varying VMA(∞) model. Section 2.4 establishes an estimation theory for the time-varying coefficients involved in a class of time-varying VARX models. Section 2.5 evaluates the finite sample performance of the proposed methods through extensive simulated and real data studies. Section 2.6 gives a short conclusion. I present some preliminary lemmas in Appendix A.1 and the proofs of the main results are given in Appendix A.2, while the proofs of the preliminary lemmas are given in Appendix A.3.

2.2 The Setup with Asymptotics

Consider the following time-varying VMA(∞) model:

$$\mathbf{x}_t = \boldsymbol{\mu}_t + \sum_{j=0}^{\infty} \mathbf{B}_{j,t}\boldsymbol{\varepsilon}_{t-j} := \boldsymbol{\mu}_t + \mathbb{B}_t(L)\boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{2.2.1}$$

where $\mathbf{x}_t$ is a vector of d-dimensional observable variables, $\boldsymbol{\mu}_t$ is a vector of d-dimensional unknown deterministic trending functions, $\mathbf{B}_{j,t}$'s are $d \times d$ unknown matrices, $\boldsymbol{\varepsilon}_t$ is a vector of d-dimensional random innovations, and d is fixed. Obviously, $\mathbb{B}_t(L) = \sum_{j=0}^{\infty} \mathbf{B}_{j,t} L^j$, where L is the lag operator. Throughout this chapter, I impose the following necessary conditions to establish the asymptotic properties.

Assumption 2.2.1. $\max_{t\geq 1} \sum_{j=1}^{\infty} j|\mathbf{B}_{j,t}| < \infty$, $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j|\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}| < \infty$, and $\limsup_{T\to\infty} \sum_{t=1}^{T-1} |\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t| < \infty$.

Assumption 2.2.2. $\{\boldsymbol{\varepsilon}_t\}_{t=-\infty}^{\infty}$ is a martingale difference sequences (m.d.s.) adapted to the filtration $\{\mathscr{F}_t\}$, where $\mathscr{F}_t = \sigma(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$ is the σ-field generated by $(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$, $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top | \mathscr{F}_{t-1}] = \mathbf{I}_d$ almost surely (a.s.), and $\max_{t\geq 1} \|\boldsymbol{\varepsilon}_t\|_\delta < \infty$ for some $\delta \geq 4$.

Assumption 2.2.1 regulates the matrices $\mathbf{B}_{j,t}$'s, and ensures the validity of the BN decomposition under a time-varying framework. It covers many cases, including (i) the parametric setting of Phillips and Solo [1992], and (ii) $\mathbf{B}_{j,t} := \mathbf{B}_j(\tau_t)$, where $\tau_t := t/T$ and $\mathbf{B}_j(\cdot)$ satisfies Lipschitz continuity on $[0,1]$ for all j. These conditions can be easily verified as they are directly related to some commonly used data generating mechanisms, see, for example, Proposition 2.2.1 below. Assumption 2.2.2 imposes conditions on the innovation error terms by replacing the commonly used independent and identically distributed (i.i.d.) innovations (e.g., Dahlhaus and Polonik, 2009) with a martingale difference structure.

I am now ready to comment on the usefulness of (2.2.1). An application of the BN decomposition to (2.2.1) immediately yields:

$$\mathbf{x}_t = \boldsymbol{\mu}_t + \mathbb{B}_t(1)\boldsymbol{\varepsilon}_t + \widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_{t-1} - \widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_t, \tag{2.2.2}$$

where $\mathbb{B}_t(L) = \mathbb{B}_t(1) - (1-L)\widetilde{\mathbb{B}}_t(L)$, $\widetilde{\mathbb{B}}_t(L) = \sum_{j=0}^{\infty} \widetilde{\mathbf{B}}_{j,t} L^j$, and $\widetilde{\mathbf{B}}_{j,t} = \sum_{k=j+1}^{\infty} \mathbf{B}_{k,t}$. The decomposition of (2.2.2) allows one to derive asymptotic properties associated with $\mathbf{x}_t$'s. For example, the following lemma holds under Assumptions 2.2.1 and 2.2.2.

Lemma 2.2.1. Under Assumptions 2.2.1-2.2.2, as $T \to \infty$, for $\forall r \in [0,1]$

$$\frac{1}{\sqrt{T}} \boldsymbol{\Sigma}^{-1/2}(r) \sum_{t=1}^{\lfloor Tr \rfloor} (\mathbf{x}_t - \boldsymbol{\mu}_t) \to_D \mathbf{W}(r),$$

where $\mathbf{\Sigma}(r) := \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{\lfloor Tr \rfloor} \mathbb{B}_t(1)\mathbb{B}_t^\top(1)$ and $\mathbf{W}(\cdot)$ is a standard multivariate Brownian motion.

By Lemma 2.2.1, it is easy to see that (2.2.1) extends similar treatments by Phillips and Solo [1992] for the univariate linear process case, and allows one to relax many $I(0)$ and $I(1)$ related results of the literature using a time-varying VMA(∞) framework.

Below, I list several examples, of which the parametric counterparts can be found in classic textbooks (e.g., Lütkepohl, 2005).

Example 1. Suppose that $\mathbf{x}_t$ is a d-dimensional time-varying VAR(p) process:

$$\mathbf{x}_t = \mathbf{A}_{1,t}\mathbf{x}_{t-1} + \cdots + \mathbf{A}_{p,t}\mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{2.2.3}$$

which has been widely studied in the literature with Bayesian framework being the dominant approach (e.g., Benati and Surico, 2009, Paul, 2020). Similar to Hamilton [1994, p.260], model (2.2.3) can be expressed as a time-varying VMA(∞) process $\mathbf{x}_t = \sum_{j=0}^{\infty} \mathbf{B}_{j,t}\boldsymbol{\varepsilon}_{t-j}$, where $\mathbf{B}_{0,t} = \mathbf{I}_d$, $\mathbf{B}_{j,t} = \mathbf{J}\prod_{i=0}^{j-1}\mathbf{\Phi}_{t-i}\mathbf{J}^\top$ for $j \geq 1$, $\mathbf{J} = [\mathbf{I}_d, \mathbf{0}_{d\times d(p-1)}]$ and $\mathbf{\Phi}_t$ is the companion matrix.

Example 2. Suppose that $\mathbf{x}_t$ is a d-dimensional time-varying VARMA(p,q) process as follows:

$$\mathbf{x}_t = \mathbf{A}_{1,t}\mathbf{x}_{t-1} + \cdots + \mathbf{A}_{p,t}\mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_{1,t}\boldsymbol{\varepsilon}_{t-1} + \cdots + \mathbf{\Theta}_{q,t}\boldsymbol{\varepsilon}_{t-q}. \tag{2.2.4}$$

Simple algebra shows that model (2.2.4) can be expressed as $\mathbf{x}_t = \sum_{b=0}^{\infty}\mathbf{D}_{b,t}\boldsymbol{\varepsilon}_{t-b}$ with $\mathbf{D}_{b,t} = \sum_{j=\max(0,b-q)}^{b}\mathbf{B}_{j,t}\mathbf{\Theta}_{b-j,t-j}$, in which $\mathbf{B}_{j,t}$ is defined similarly as in Example 1, and $\mathbf{\Theta}_{0,t} = \mathbf{I}_d$ is independent of t.

Example 3. Suppose that $\mathbf{x}_t$ is a d-dimensional time-varying VARX process of the form:

$$\mathbf{x}_t = \mathbf{A}_{1,t}\mathbf{x}_{t-1} + \ldots + \mathbf{A}_{p,t}\mathbf{x}_{t-p} + \mathbf{\Theta}_t\mathbf{z}_t + \boldsymbol{\varepsilon}_t \quad \text{and} \quad \mathbf{z}_t = \sum_{j=0}^{\infty}\mathbf{C}_{j,t}\mathbf{v}_{t-j}, \tag{2.2.5}$$

where $\mathbf{z}_t$ is an m-dimensional vector and $\mathbf{\Theta}_t$ is a $d \times m$ matrix. Then model (2.2.5) can be further written as

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} = \sum_{j=0}^{\infty} \begin{bmatrix} \mathbf{B}_{j,t} & \mathbf{D}_{j,t} \\ \mathbf{0} & \mathbf{C}_{j,t} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{t-j} \\ \mathbf{v}_{t-j} \end{bmatrix},$$

where $\mathbf{D}_{j,t} = \sum_{k=0}^{j} \mathbf{B}_{k,t}\boldsymbol{\Theta}_{t-k}\mathbf{C}_{j-k,t-k}$ and $\mathbf{B}_{j,t}$ is defined in a way similar to that in Example 1.

The following proposition shows that each of the models listed in the above three examples may have a time-varying VMA(∞) representation.

Proposition 2.2.1.

1. Consider Examples 1 and 2. Suppose that the roots of $\mathbf{I}_d - \mathbf{A}_{1,t} - \cdots - \mathbf{A}_{p,t} = \mathbf{0}_d$ all lie outside the unit circle uniformly over t, $\limsup_{T\to\infty} \sum_{t=1}^{T-1} |\mathbf{A}_{m,t+1} - \mathbf{A}_{m,t}| < \infty$ for $m = 1, \ldots, p$ and $\mathbf{A}_{m,t} = \mathbf{A}_{m,1}$ for $t \leq 0$ and $m = 1, \ldots, p$. In addition, suppose that in Example 2, $\limsup_{T\to\infty} \sum_{t=1}^{T-1} |\boldsymbol{\Theta}_{m,t+1} - \boldsymbol{\Theta}_{m,t}| < \infty$ for $m = 1, \ldots, q$. Then both (2.2.3) and (2.2.4) are time-varying VMA(∞) processes, for which the coefficients satisfy Assumption 2.2.1.

2. Suppose that $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j|\mathbf{C}_{j,t+1} - \mathbf{C}_{j,t}| < \infty$ and $\max_{t\geq 1} \sum_{j=1}^{\infty} j|\mathbf{C}_{j,t}| < \infty$ for Example 3. Moreover, let $\mathbf{A}_{j,t}$ and $\boldsymbol{\Theta}_t$ satisfy the same conditions as those in the first result of this proposition. Then (2.2.5) is a time-varying VMA(∞), for which the coefficients satisfy Assumption 2.2.1.

Estimation and testing issues for the coefficient matrices involved in Examples 1-3 and their semiparametric counterparts are of significant interest in econometrics, and should be fully investigated in separate research. Section 2.4 below considers non- and semi-parametric estimation problems for Example 3.

I next explain the advantages of this VMA setting compared to locally stationary and mixing conditions below. Let's assume $x_{t,T}$ is a locally stationary process such that for each rescaled time point $\tau \in [0,1]$ there exists an approximated stationary process $\{x_t(\tau)\}_{t=1}^{T}$ satisfying

$$\sup_{t\geq 1}\{E\|x_{t,T} - x_t(\tau_t)\|^{\delta}\}^{1/\delta} = O(1/T) \text{ and } \sup_{\tau,\tau'}\{E\|x_t(\tau) - x_t(\tau')\|^{\delta}\}^{1/\delta} = O(|\tau - \tau'|) \quad (2.2.6)$$

which is originally proposed by Dahlhaus and Rao [2006] in the context of time–varying ARCH models. Assuming $Th^3 \to 0$ (that is in fact unusual in nonparametric regressions), the stationary approximation error becomes negligible, and Dahlhaus and Rao [2006] establish the central limit theory for local constant QML estimator of time–varying ARCH models.

Having (2.2.6) in hand, one then needs to define a proper dependence measure for $\{x_t(\tau)\}_{t=1}^T$ in order to establish asymptotic theories. Many works turn to locally mixing framework assuming that $\{x_t(\tau)\}_{t=1}^T$ is stationary mixing for any given $\tau \in [0,1]$ (e.g., Pei et al., 2018, Sun et al., 2021b), while some papers directly impose cumulant conditions on $\{x_t(\tau)\}_{t=1}^T$ (e.g., Assumption 3.2 in Casini, forthcoming) and argue that these cumulant conditions can be verified by using mixing condition.

In view of the aforementioned literature, I argue that the linear process is more suitable in many scenarios. For example, we consider a class of multivariate autoregressive models, including those discussed in Examples 1 and 2, in which each $\{x_t\}$ process itself is autoregressive and generated by a time–varying structure. Therefore, by Proposition 2.2.1, both examples admit VMA(∞) representations, which allow us to further establish asymptotic properties accordingly.

In contrast, it is difficult to justify that a time–varying dynamic model satisfies mixing conditions. Indeed, even stationary linear processes (including some simple AR(1) processes) are not necessarily mixing (Doukhan, 2012, Section 2.3.1), unless some restrictive conditions on densities of error terms are imposed (Withers, 1981). That said, in this chapter I adopt the linear process structure for the time–varying setting. Along with the proposed BN decomposition, I am able to establish estimation and inferential theories for a wide class of time–varying dynamic models without imposing additional cost.

2.2.1 Asymptotic Properties of the Sample Moments

In this subsection, I present some useful asymptotic properties associated with (2.2.1). First, I propose the law of large numbers for two weighted sample moments of $\mathbf{x}_t$.

Lemma 2.2.2. Let Assumptions 2.2.1 and 2.2.2 hold. In addition, suppose that $\{\mathbf{W}_{T,t}\}_{t=1}^T$ is a sequence of $m \times d$ deterministic weighting matrices satisfying (1) $\sum_{t=1}^T |\mathbf{W}_{T,t}| =$

$O(1)$, (2) $\max_{t\geq 1} |\mathbf{W}_{T,t}| = O(d_T)$, and (3) $\sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1} - \mathbf{W}_{T,t}| = O(d_T)$, where $d_T = \max_{t\geq 1} |\mathbf{W}_{T,t}| \to 0$. Then, as $T \to \infty$,

$$\sum_{t=1}^{T} \mathbf{W}_{T,t} \left(\mathbf{x}_t - E(\mathbf{x}_t)\right) = O_P(\sqrt{d_T}) \quad \text{and} \quad \sum_{t=1}^{T} \mathbf{W}_{T,t} \left(\mathbf{x}_t \mathbf{x}_{t+p}^\top - E\left(\mathbf{x}_t \mathbf{x}_{t+p}^\top\right)\right) = O_P(\sqrt{d_T}),$$

where both m (≥ 1) and p (≥ 0) are fixed integers.

Lemma 2.2.2 provides the conditions that ensure the convergence of weighted sample moments, and can be easily applied to study weighted-least-squares type estimators. It is worth stressing that the conditions on $\{\mathbf{W}_{T,t}\}$ are weak, in the sense that Lemma 2.2.2 covers both the parametric rate $d_T = \frac{1}{T}$ (e.g., $\mathbf{W}_{T,t} = \frac{1}{T}$) and the nonparametric rate $d_T = \frac{1}{Th}$ (e.g., $\mathbf{W}_{T,t} = \frac{1}{T} K_h(\tau_t - \tau)$ for $\forall \tau \in [0,1]$).

Next, I strengthen the results of Lemma 2.2.2, and establish the rates of uniform convergence.

Lemma 2.2.3. Let Assumptions 2.2.1 and 2.2.2 hold. In addition, let $\{\mathbf{W}_{T,t}(\cdot)\}_{t=1}^{T}$ be a sequence of $m \times d$ matrices of deterministic weighting functions, in which m is fixed, and each functional component is Lipschitz continuous and defined on a compact set $[a,b]$. Moreover, suppose that

1. $\sup_{\tau\in[a,b]} \sum_{t=1}^{T} |\mathbf{W}_{T,t}(\tau)| = O(1)$;
2. $\sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)| = O(d_T)$, where $d_T = \sup_{\tau\in[a,b],t\geq 1} |\mathbf{W}_{T,t}(\tau)| \to 0$.

Then as $T \to \infty$,

1. $\sup_{\tau\in[a,b]} \left| \sum_{t=1}^{T} \mathbf{W}_{T,t}(\tau) \left(\mathbf{x}_t - E(\mathbf{x}_t)\right) \right| = O_P(\sqrt{d_T \log T})$ provided $T^{\frac{2}{\delta}} d_T \log T \to 0$;
2. $\sup_{\tau\in[a,b]} \left| \sum_{t=1}^{T} \mathbf{W}_{T,t}(\tau) \left(\mathbf{x}_t \mathbf{x}_{t+p}^\top - E(\mathbf{x}_t \mathbf{x}_{t+p}^\top)\right) \right| = O_P(\sqrt{d_T \log T})$ for any fixed integer $p \geq 0$ provided $T^{\frac{4}{\delta}} d_T \log T \to 0$ and $\max_{t\geq 1} E(|\boldsymbol{\varepsilon}_t|^4 | \mathscr{F}_{t-1}) < \infty$ a.s., where $\delta > 4$ is the same as that in Assumption 2.2.2.

Lemma 2.2.3 corresponds to some existing uniform convergence results for nonparametric estimation of many time series models associated with either stationarity or unit roots, such as those in Hansen [2008], Gao et al. [2015], Li et al. [2016], and Phillips et al. [2017].

As a specific application, in Section 2.4 I apply this result to establish an estimation theory for a class of nonparametric and semiparametric time-varying VARX models.

2.2.2 Inferences

To obtain valid inferences in practice, in what follows I establish a central limit theory in Lemma 2.2.4, and then propose two methods (i.e., the dependent wild bootstrap (DWB) approach and the heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimation approach) to estimate the asymptotic covariance matrix in Lemmas 2.2.5 and 2.2.6, respectively.

Lemma 2.2.4. Let Assumptions 2.2.1-2.2.2 hold. Suppose $\{\mathbf{W}_{T,t}\}_{t=1}^{T}$ is a sequence $m \times d$ deterministic weighting matrices satisfying (1) $\sum_{t=1}^{T} |\mathbf{W}_{T,t}| = O(1)$, (2) $\max_{t\geq 1} |\mathbf{W}_{T,t}| = O(d_T)$ and (3) $\sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1} - \mathbf{W}_{T,t}| = O(d_T)$, where the sequence of real numbers, d_T, is chosen to ensure that $\mathbf{\Sigma_W} = \lim_{T\to\infty} \frac{1}{d_T} \sum_{t=1}^{T} \mathbf{W}_{T,t}\mathbb{B}_t(1)\mathbb{B}_t^\top(1)\mathbf{W}_{T,t}^\top$ is a positive definite matrix. As $T \to \infty$, I then have

$$\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \mathbf{W}_{T,t} \left(\mathbf{x}_t - E(\mathbf{x}_t)\right) \to_D N(\mathbf{0}, \mathbf{\Sigma_W}),$$

where m (≥ 1) is a fixed positive integer.

With Lemma 2.2.4 in hand, in order to infer the population mean of $\mathbf{x}_t$, the only missing piece is the information of $\mathbf{\Sigma_W}$, which is a type of long-run covariance matrix arising from the infinity memory of $\mathbf{x}_t$. To recover $\mathbf{\Sigma_W}$, I then consider two approaches: (i) the DWB approach, and (ii) the HAC covariance matrix estimation. Both approaches date back to Shao [2010], and Newey and West [1987], respectively.

I start with the DWB method, and suppose that $\{\xi_t^*\}_{t=1}^{T}$ is a sequence of l-dependent time series satisfying $E[\xi_t^*] = 0$, $E[\xi_t^{*2}] = 1$, $E|\xi_t^*|^\delta < \infty$ for some $\delta > 2$, and $E[\xi_t^*\xi_s^*] = a\left((t-s)/l\right)$ for a kernel function $a(\cdot)$ and a tuning parameter l. The DWB procedure requires a tuning parameter l, which is the "block length" (Shao, 2010) ensuing the variables further than l units apart are independent.

Lemma 2.2.5. Let $l \to \infty$ and $l\sqrt{d_T} \to 0$. Additionally, let $a(\cdot)$ be a symmetric kernel and Lipschitz continuous on $[-1, 1]$ satisfying that $a(0) = 1$ and

$$K_a(x) = \int_{-\infty}^{\infty} a(u)e^{-iux}\mathrm{d}u \geq 0 \quad \text{for } x \in \mathbb{R}.$$

Under the conditions of Lemma 2.2.4, as $T \to \infty$,

$$\sup_{\boldsymbol{w}\in\mathbb{R}^d} \left| \Pr{}^* \left[\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \widetilde{\mathbf{x}}_t \xi_t^* \leq \boldsymbol{w} \right] - \Pr\left[\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \widetilde{\mathbf{x}}_t \leq \boldsymbol{w} \right] \right| = o_P(1),$$

where $\widetilde{\mathbf{x}}_t = \mathbf{W}_{T,t}(\mathbf{x}_t - E(\mathbf{x}_t))$, and $\Pr^*$ denotes the probability measure induced by the DWB procedure.

Lemma 2.2.5 establishes the consistency of the bootstrap version of the weighted sample mean of $\mathbf{x}_t$. The condition of $K_a(x)$ ensures the semi-positive definiteness of the covariance matrix of $\{\xi_t^*\}_{t=1}^T$, while the restrictions on $a(\cdot)$ are satisfied by a few commonly used kernels, such as the Bartlett and Parzen kernels.

I now consider the HAC approach to deal with inferential issues. Specifically, I define

$$\widehat{\mathbf{\Sigma}}_{\mathbf{W}} = \widehat{\mathbf{\Xi}}_0 + \sum_{i=1}^{b} \psi\left(i/b\right)(\widehat{\mathbf{\Xi}}_i + \widehat{\mathbf{\Xi}}_i^\top), \tag{2.2.7}$$

where $\widehat{\mathbf{\Xi}}_i(\tau) = \frac{1}{d_T}\sum_{t=1}^{T-i} \mathbf{W}_{T,t}\mathbf{e}_t\mathbf{e}_{t+i}^\top\mathbf{W}_{T,t+i}^\top$ for $i \geq 0$, $\mathbf{e}_t = \mathbf{x}_t - E(\mathbf{x}_t)$, $\psi(\cdot)$ is a kernel function, and b is the bandwidth diverging at a relatively slow rate, in which $E(\mathbf{x}_t)$ is assumed to be computable at this stage. Otherwise, it will be replaced by an estimated version as in (2.3.3) below. Under some mild conditions, I establish asymptotic properties for (2.2.7) in the following lemma.

Lemma 2.2.6. Suppose that $\psi(\cdot)$ is Lipschitz continuous, and has a compact support on $[-1, 1]$ with $\psi(0) = 1$. Additionally, let $b \to \infty$ and $b\sqrt{d_T} \to 0$. Under the conditions of Lemma 2.2.4, $\widehat{\mathbf{\Sigma}}_{\mathbf{W}} = \mathbf{\Sigma}_{\mathbf{W}} + o_P(1)$.

The conditions on b and $\psi(\cdot)$ are standard, and are similar to those for the DWB method. For the case of parametric estimation ($d_T = \frac{1}{T}$), the condition $b\sqrt{d_T} \to 0$ is identical to that of Hansen [1992], who proves the consistency of long-run covariance matrix estimator under the mixing condition. Apart from constructing confidence intervals for weighted sample mean, the long-run covariance estimation is also essential in model specification testing, see, for example, Zhang and Wu [2011].

Up to this point, I have established a set of asymptotic properties for the VMA(∞) process (2.2.1). In the following section, I apply these results to study the smooth deterministic trends of (2.2.1) for the purpose of estimation and inference.

2.3 On the Deterministic Trends

To facilitate the development, it is useful to impose the following specifications:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}(\tau_t) \quad \text{and} \quad \mathbf{B}_{j,t} = \mathbf{B}_j(\tau_t),$$

where $\tau_t = t/T$. Thus, (2.2.1) can be rewritten as

$$\mathbf{x}_t = \boldsymbol{\mu}(\tau_t) + \sum_{j=0}^{\infty} \mathbf{B}_j(\tau_t)\boldsymbol{\varepsilon}_{t-j}. \tag{2.3.1}$$

Below I show that the trending function $\boldsymbol{\mu}(\tau)$ and the long-run covariance matrix associated with (2.3.1) can be well recovered, although I am unable to consistently estimate each individual $\mathbf{B}_j(\tau)$. The following assumptions are necessary.

Assumption 2.3.1. Each component of $\boldsymbol{\mu}(\cdot)$ and $\mathbf{B}_j(\cdot)$'s is second order continuously differentiable on $[0,1]$. Moreover, $\sup_{\tau\in[0,1]} \sum_{j=1}^{\infty} j|\mathbf{B}_j^{(\ell)}(\tau)| < \infty$ for $\ell = 0, 1$.

Assumption 2.3.2. 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 $T \to \infty$, $h \to 0$ and $Th \to \infty$.

Assumption 2.3.1 imposes certain smoothness conditions on the functional coefficients, which are easily verifiable and can be regarded as a special case of Assumption 2.2.1. Assumption 2.3.2 is standard in the literature of nonparametric kernel estimation (cf., Li and Racine, 2007).

With these conditions in hand, I estimate $\boldsymbol{\mu}(\tau)$ by

$$\widehat{\boldsymbol{\mu}}(\tau) = \left[\sum_{t=1}^{T} K_h(\tau_t - \tau)\right]^{-1} \sum_{t=1}^{T} \mathbf{x}_t K_h(\tau_t - \tau), \tag{2.3.2}$$

and establish an asymptotic distribution in Theorem 2.3.1.

Theorem 2.3.1. Let Assumptions 2.2.2-2.3.2 hold. If, in addition, $Th^5 \to \alpha \in [0, \infty)$, then, for $\forall \tau \in (0,1)$, as $T \to \infty$,

$$\sqrt{Th}(\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{\mu}(\tau)) \to_D N\left(\boldsymbol{\mu}_b(\tau), \tilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)\right),$$

where $\boldsymbol{\mu}_b(\tau) = \frac{1}{2}\sqrt{\alpha}\tilde{c}_2 \boldsymbol{\mu}^{(2)}(\tau)$, $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau) = \sum_{j=0}^{\infty} \mathbf{B}_j(\tau) \sum_{j=0}^{\infty} \mathbf{B}_j^\top(\tau)$, and $\tilde{c}_2$ and $\tilde{v}_0$ have been defined in Section 2.1.

If $\alpha = 0$, there is no bias term involved in the asymptotic distribution of Theorem 2.3.1, which then falls in the usual undersmoothing scenario (Li and Racine, 2007). To establish valid inferences, both $\boldsymbol{\mu}_b(\tau)$ and $\tilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$ have to be estimated, and I apply both the DWB and HAC methods of Section 2.2. In particular, the DWB procedure is able to handle the estimation of both $\boldsymbol{\mu}_b(\tau)$ and $\tilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$ simultaneously.

The DWB Method – The implementation is as follows:

1. For $\forall \tau \in (0,1)$, let $\widetilde{\boldsymbol{\mu}}(\tau)$ be defined in the same way as in (2.3.2) using an oversmoothing bandwidth $\widetilde{h}$, and obtain the residuals: $\widetilde{\mathbf{e}}_t = \mathbf{x}_t - \widetilde{\boldsymbol{\mu}}(\tau_t)$ for $t \geq 1$.

2. Generate $\mathbf{x}_t^* = \widetilde{\boldsymbol{\mu}}(\tau_t) + \mathbf{e}_t^*$ with $\mathbf{e}_t^* = \xi_t^* \widetilde{\mathbf{e}}_t$, in which ξ_t^*'s form an l-dependent time series satisfying $E[\xi_t^*] = 0$, $E[\xi_t^{*2}] = 1$, $E|\xi_t^*|^\delta < \infty$ for some $\delta > 2$, and $E[\xi_t^* \xi_s^*] = a\left((t-s)/l\right)$ with a kernel function $a(\cdot)$ and a tuning parameter l.

3. Use $\mathbf{x}_t^*$'s to construct an estimator $\widehat{\boldsymbol{\mu}}^*(\tau)$ as in (2.3.2).

4. Repeat Steps 2-3 J times. Let $\boldsymbol{q}_\alpha(\tau)$ be the α-quantile of the J statistics $\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)$, and denote the $(1-\alpha)\cdot 100\%$ confidence interval of $\widehat{\boldsymbol{\mu}}(\tau)$ as

$$\left[\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{1-\alpha/2}(\tau),\ \widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{\alpha/2}(\tau)\right].$$

Here, $\widetilde{h}$ is an oversmoothing bandwidth, as I shall require $h/\widetilde{h} \to 0$, where h is the same as that in (2.3.2). The asymptotic properties for the DWB procedure are given in Theorem 2.3.2 below.

Theorem 2.3.2. Let $l \to \infty$, $\max\{\widetilde{h}, h/\widetilde{h}\} \to 0$ and $l \cdot \max\{1/\sqrt{Th}, \widetilde{h}^4\} \to 0$. Additionally, let $a(\cdot)$ be a symmetric kernel and Lipschitz continuous on $[-1,1]$ satisfying that $a(0) = 1$ and

$$K_a(x) = \int_{-\infty}^{\infty} a(u)e^{-iux}\mathrm{d}u \geq 0 \quad \text{for } x \in \mathbb{R}.$$

Under the conditions of Theorem 2.3.1, for $\forall \tau \in (0,1)$

1. $\sup_{\boldsymbol{w}\in\mathbb{R}^d} \left| \Pr^* \left[\sqrt{Th}\,(\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)) \leq \boldsymbol{w} \right] - \Pr\left[\sqrt{Th}\,(\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{\mu}(\tau)) \leq \boldsymbol{w} \right] \right| = o_P(1)$,
2. $\liminf_{T\to\infty} \Pr\left(\boldsymbol{\mu}(\tau) \in \left[\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{1-\alpha/2}(\tau), \widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{\alpha/2}(\tau)\right]\right) = 1 - \alpha$,

where $\Pr^*$ denotes the probability measure induced by the DWB procedure.

Theorem 2.3.2 shows that the confidence interval of $\boldsymbol{\mu}(\tau)$ can be recovered by the empirical quantile of $\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)$. Note that there is no need to deal with the bias in the DWB procedure, as the bootstrap draws generate a bias term identical to that in Theorem 2.3.1 (see (A.2.5) of Appendix A for the technical details).

The HAC Method – The HAC estimator is naturally given by:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) = \widehat{\boldsymbol{\Xi}}_0(\tau) + \sum_{i=1}^{b} \psi\left(i/b\right)(\widehat{\boldsymbol{\Xi}}_i(\tau) + \widehat{\boldsymbol{\Xi}}_i^\top(\tau)), \tag{2.3.3}$$

where $\widehat{\boldsymbol{\Xi}}_i(\tau) = \left[\sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right)\right]^{-1} \sum_{t=1}^{T-i} \widehat{\mathbf{e}}_t \widehat{\mathbf{e}}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)$ for $i \geq 0$, $\widehat{\mathbf{e}}_t = \mathbf{x}_t - \widehat{\boldsymbol{\mu}}(\tau_t)$, $\psi(\cdot)$ is a kernel function, and b is the bandwidth diverging at a relatively slow rate.

Under some mild conditions, I summarize the asymptotic property of (2.3.3) in the following theorem.

Theorem 2.3.3. Suppose that $\psi(\cdot)$ is Lipschitz continuous, and has a compact support on $[-1,1]$ with $\psi(0) = 1$. Additionally, let $b \to \infty$ and $b/\sqrt{Th} \to 0$. Under the conditions of Theorem 2.3.1, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) = \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau) + o_P(1)$ for $\forall \tau \in (0,1)$.

It should be pointed out that the HAC method does not handle the bias term at all, so it only generates valid inference when $Th^5 \to 0$. To estimate the bias term in this case, one will have to employ a higher-order local polynomial approach as in Xia [1998] and Hall and Racine [2015]. I no longer pursue the latter in this study.

As a further application, in the next section I establish uniform consistency of nonparametric kernel estimators of the time-varying coefficients in a class of time-varying VARX models. In addition, one could estimate the parametric components with a $\sqrt{T}$-convergence rate for semiparametric time-varying VARX models.

2.4 Estimation of Time-Varying VARX Models

In this section I use the results in Section 2.2 to derive asymptotic properties for non- and semi-parametric estimators in a class of time-varying VARX models of the form:

$$\mathbf{y}_t = \sum_{j=1}^{p} \mathbf{A}_j(\tau_t)\mathbf{y}_{t-j} + \sum_{j=0}^{q} \mathbf{B}_j(\tau_t)\mathbf{x}_{t-j} + \boldsymbol{\eta}_t := \mathbf{Z}_t^\top \boldsymbol{\beta}(\tau_t) + \boldsymbol{\eta}_t \tag{2.4.1}$$

where $\mathbf{Z}_t = \mathbf{z}_t \otimes \mathbf{I}_d$, $\mathbf{z}_t = (\mathbf{y}_{t-1}^\top, \ldots, \mathbf{y}_{t-p}^\top, \mathbf{x}_t^\top, \mathbf{x}_{t-1}^\top, \ldots, \mathbf{x}_{t-q}^\top)^\top$, and $\boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\varepsilon}_t$. Here, $\mathbf{y}_t = (y_{1,t}, \ldots, y_{d,t})^\top$ is a d-dimensional vector of endogenous variables, $\mathbf{x}_t = (x_{1,t}, \ldots, x_{m,t})^\top$ is an m-dimensional vector of exogenous variables, and both d and m are finite integers. Accordingly, $\{\mathbf{A}_j(\tau)\}$ and $\{\mathbf{B}_j(\tau)\}$ are the $d \times d$ and $d \times m$ coefficient matrices. Also, $\boldsymbol{\omega}(\tau)$ is an unknown deterministic function which has full row rank uniformly in $\tau \in [0, 1]$, and captures the heteroscedasticity over time. Obviously, I have

$$\boldsymbol{\beta}(\tau) = \mathrm{vec}(\mathbf{A}(\tau), \mathbf{B}(\tau)), \tag{2.4.2}$$

where $\mathbf{A}(\tau) = (\mathbf{A}_1(\tau), \ldots, \mathbf{A}_p(\tau))$ and $\mathbf{B}(\tau) = (\mathbf{B}_0(\tau), \mathbf{B}_1(\tau), \ldots, \mathbf{B}_q(\tau))$.

In what follows, I am interested to estimate $\{\mathbf{A}_j(\tau)\}$ and $\{\mathbf{B}_j(\tau)\}$, and are particularly interested to adopt the nonparametric local linear approach and the semiparametric profile likelihood estimation of Fan and Huang [2005]. In the following two subsections, I consider both non- and semi-parametric versions of time-varying VARX models.

2.4.1 Nonparametric Estimation

I start with $\boldsymbol{\beta}(\cdot)$, and assume each component of $\boldsymbol{\beta}(\cdot)$ has continuous derivatives up to the second order. When τ_t is close to τ, I then have the following approximation:

$$\mathbf{y}_t \simeq \mathbf{Z}_t^\top \boldsymbol{\beta}(\tau) + \mathbf{Z}_t^\top \boldsymbol{\beta}^{(1)}(\tau)(\tau_t - \tau) + \boldsymbol{\eta}_t. \tag{2.4.3}$$

Usually, $\{\boldsymbol{\beta}(\tau), \boldsymbol{\beta}^{(1)}(\tau)\}$ of (2.4.3) can be estimated by the kernel weighted least-squares criterion:

$$(\widehat{\boldsymbol{\beta}}(\tau), \widehat{\boldsymbol{\beta}}^{(1)}(\tau)) = \underset{\boldsymbol{\beta}, \boldsymbol{\beta}^{(1)}}{\operatorname{argmin}} \sum_{t=1}^{T} \left| \mathbf{y}_t - \mathbf{Z}_t^\top (\boldsymbol{\beta} + (\tau_t - \tau)\boldsymbol{\beta}^{(1)}) \right|^2 K_h(\tau_t - \tau). \tag{2.4.4}$$

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

$$\widehat{\boldsymbol{\beta}}(\tau) = (\mathbf{I}_l, \mathbf{0}_l)(\mathbf{Z}_\tau^\top \boldsymbol{K}_\tau \mathbf{Z}_\tau)^{-1} \mathbf{Z}_\tau^\top \boldsymbol{K}_\tau \mathbf{y}, \tag{2.4.5}$$

where $l = d^2 p + (q+1)md$, $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_T^\top)^\top$,

$$\boldsymbol{K}_\tau = \operatorname{diag}\{K_h(\tau_1 - \tau), \ldots, K_h(\tau_T - \tau)\} \otimes \mathbf{I}_d, \quad \text{and} \quad \mathbf{Z}_\tau = \begin{pmatrix} \mathbf{Z}_1^\top & \mathbf{Z}_1^\top \frac{\tau_1 - \tau}{h} \\ \vdots & \vdots \\ \mathbf{Z}_T^\top & \mathbf{Z}_T^\top \frac{\tau_T - \tau}{h} \end{pmatrix}.$$

Having presented the above estimators, I introduce the following assumptions for the theoretical development.

Assumption 2.4.1.

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

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

3. Suppose that $\mathbf{x}_t = \mathbf{g}(\tau_t) + \sum_{j=0}^{\infty} \mathbf{D}_j(\tau_t)\mathbf{v}_{t-j}$ for $t \geq 1$ and $\mathbf{x}_t = \mathbf{g}(0) + \sum_{j=0}^{\infty} \mathbf{D}_j(0)\mathbf{v}_{t-j}$ for $t \leq 0$, where $\mathbf{g}(\cdot)$ and $\mathbf{D}_j(\cdot)$ are $m \times 1$ and $m \times m$ respectively. Each component

of $\mathbf{g}(\cdot)$ and $\mathbf{D}_j(\cdot)$ is second-order continuously differentiable on $[0,1]$. For $\ell = 0,1$, $\sup_{\tau\in[0,1]}\sum_{j=1}^{\infty} j|\mathbf{D}_j^{(\ell)}(\tau)| < \infty$.

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

5. Let $\mathbf{e}_t = (\boldsymbol{\varepsilon}_t^\top, \mathbf{v}_{t+1}^\top)^\top$ and $\{\mathbf{e}_t\}_{t=-\infty}^{\infty}$ form a sequence of martingale differences such that $E\left(\mathbf{e}_t|\mathscr{F}_{t-1}\right) = 0$, where $\mathscr{F}_t = \sigma\{\mathbf{e}_t, \mathbf{e}_{t-1}, \ldots\}$. Also, suppose that $E(\mathbf{e}_t\mathbf{e}_t^\top|\mathscr{F}_{t-1}) = \begin{pmatrix} \mathbf{I}_d & \boldsymbol{\rho}_{\boldsymbol{\varepsilon}\mathbf{v}} \\ \boldsymbol{\rho}_{\boldsymbol{\varepsilon}\mathbf{v}}^\top & \mathbf{I}_m \end{pmatrix}$ a.s., and $\max_{t\geq 1}\|\mathbf{e}_t\|_\delta < \infty$ for some $\delta \geq 4$.

Assumptions 2.4.1.2 is pretty standard in the literature (Li and Racine, 2007), so the discussions are omitted. Assumption 2.4.1.5 is also standard by assuming that the innovation errors follow a martingale difference structure, which is identical to those used in Phillips and Lee [2013] for example.

I now comment on the rest conditions of Assumption 2.4.1. Assumption 2.4.1.1 ensures that $\mathbf{y}_t$ in model (2.4.1) is neither a unit-root process nor an explosive process, and can be regarded as an extension of those used for the classical multivariate dynamic models (e.g., Hamilton, 1994, p. 259). Assumption 2.4.1.3 formulates a time-varying VMA(∞) process which nests many different processes as special cases as shown in Examples 1-3. Assumption 2.4.1.4 imposes the heteroscedasticity on the structure using an unknown function, and the assumptions are in the same spirit of those for $\boldsymbol{\beta}(\cdot)$.

The following theorem establishes the asymptotic properties associated with the estimation procedure of (2.4.5).

Theorem 2.4.1. Let Assumptions 2.3.2 and 2.4.1 hold. If $T \to \infty$, then

1. $\forall \tau \in (0,1)$ I have

$$\sqrt{Th}(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) - \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\beta}^{(2)}(\tau) + o_P(h^2)) \to_D N\left(\mathbf{0}, \widetilde{v}_0\mathbf{V}(\tau)\right),$$

where $\mathbf{V}(\tau) = \boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\tau)\otimes\boldsymbol{\Omega}(\tau)$ and $\boldsymbol{\Sigma}_{\mathbf{z}}(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E(\mathbf{z}_t\mathbf{z}_t^\top)K_h(\tau_t - \tau)$;

If, in addition, $\max_{t\geq 1} E\left[|\mathbf{e}_t|^4 |\mathscr{F}_{t-1}\right] < \infty$ a.s. and $\frac{T^{1-4/\delta}h}{\log T} \to \infty$, then I have

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

3. $\widehat{\mathbf{V}}(\tau) \to_P \mathbf{V}(\tau)$ for $\forall \tau \in [0,1]$, where $\widehat{\mathbf{V}}(\tau) = \widehat{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}(\tau)\otimes\widehat{\boldsymbol{\Omega}}(\tau)$, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{z}}(\tau) = (\frac{1}{T}\sum_{t=1}^T K_h(\tau_t-\tau))^{-1}\frac{1}{T}\sum_{t=1}^T \mathbf{z}_t\mathbf{z}_t^\top K_h(\tau_t-\tau)$, $\widehat{\boldsymbol{\Omega}}(\tau) = (\frac{1}{T}\sum_{t=1}^T K_h(\tau_t-\tau))^{-1}\frac{1}{T}\sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top K_h(\tau_t-\tau)$ and $\widehat{\boldsymbol{\eta}}_t = \mathbf{y}_t - \mathbf{Z}_t^\top\widehat{\boldsymbol{\beta}}(\tau_t)$.

2.4.2 Semiparametric Estimation

In this subsection, I consider a semiparametric version of model (2.4.1), assuming that some of $\boldsymbol{\beta}(\cdot)$ are not time-varying. Let $\mathbf{C}$ be a $s\times l$ selection matrix with $1\le s< l$ such that $\mathbf{C}\boldsymbol{\beta}(\cdot) = \mathbf{c}$. In addition, let $\widetilde{\mathbf{C}}$ be a selection matrix collecting the elements of $\boldsymbol{\beta}(\tau)$ left out by $\mathbf{C}$. Thus, (2.4.1) can be rewritten as

$$\mathbf{y}_t = \mathbf{X}_{\mathbf{C},t}^\top\mathbf{c} + \mathbf{X}_{\widetilde{\mathbf{C}},t}^\top\boldsymbol{\theta}(\tau) + \boldsymbol{\eta}_t, \tag{2.4.6}$$

where $\mathbf{X}_{\mathbf{C},t} = \mathbf{C}\mathbf{Z}_t$, $\mathbf{X}_{\widetilde{\mathbf{C}},t} = \widetilde{\mathbf{C}}\mathbf{Z}_t$, and $\boldsymbol{\theta}(\tau) = \widetilde{\mathbf{C}}\boldsymbol{\beta}(\tau)$. The right hand side of (2.4.6) reduces to a semiparametric time-varying model. Using the profile likelihood estimation therein, $\mathbf{c}$ can be estimated by

$$\widehat{\mathbf{c}} = \left(\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT}-\mathbf{S})^\top(\mathbf{I}_{dT}-\mathbf{S})\mathbf{X}_{\mathbf{C}}\right)^{-1}\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT}-\mathbf{S})^\top(\mathbf{I}_{dT}-\mathbf{S})\mathbf{y}, \tag{2.4.7}$$

where $\mathbf{S} = (\mathbf{s}(\tau_1)^\top\mathbf{X}_{\widetilde{\mathbf{C}},1},\ldots,\mathbf{s}(\tau_T)^\top\mathbf{X}_{\widetilde{\mathbf{C}},T})^\top$, $\mathbf{s}(\tau) = (\mathbf{I}_{l-s},\mathbf{0}_{l-s})(\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^\top\boldsymbol{K}_\tau\mathbf{X}_{\widetilde{\mathbf{C}},\tau})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^\top\boldsymbol{K}_\tau$, and

$$\mathbf{X}_{\widetilde{\mathbf{C}},\tau} = \begin{pmatrix} \mathbf{X}_{\widetilde{\mathbf{C}},1}^\top & \mathbf{X}_{\widetilde{\mathbf{C}},1}^\top\frac{\tau_1-\tau}{h} \\ \vdots & \vdots \\ \mathbf{X}_{\widetilde{\mathbf{C}},T}^\top & \mathbf{X}_{\widetilde{\mathbf{C}},T}^\top\frac{\tau_T-\tau}{h} \end{pmatrix}.$$

Finally, $\boldsymbol{\theta}(\tau)$ of (2.4.6) can be estimated by

$$\widehat{\boldsymbol{\theta}}(\tau) = (\mathbf{I}_{l-s},\mathbf{0}_{l-s})(\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^\top\boldsymbol{K}_\tau\mathbf{X}_{\widetilde{\mathbf{C}},\tau})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^\top\boldsymbol{K}_\tau\left(\mathbf{y}-\mathbf{X}_{\mathbf{C}}\widehat{\mathbf{c}}\right), \tag{2.4.8}$$

where $\mathbf{X}_{\mathbf{C}} = (\mathbf{X}_{\mathbf{C},1},\ldots,\mathbf{X}_{\mathbf{C},T})^\top$.

Having proposed the above estimators, I introduce following assumption for the establishment of a semiparametric estimation theory.

Assumption 2.4.2. Let $\max_{t\geq 1} E[|\mathbf{e}_t|^4 \,|\mathscr{F}_{t-1}] < \infty$ a.s., $Th^8 \to 0$, $\frac{Th^2}{(\log T)^2} \to \infty$, $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$, and $\delta > 4$, where δ is the same as that of Assumption 2.4.1.5.

Assumption 2.4.2 imposes more restrictive conditions on the bandwidth, and the conditional moments of the error terms. These assumptions are commonly used in the literature of semiparametric kernel estimation [e.g., Fan and Huang, 2005].

With Assumptions 2.4.1 and 2.4.2 in hand, the next theorem establishes the asymptotic distributions associated with the estimation procedure of (2.4.7) and (2.4.8).

Theorem 2.4.2. Let Assumptions 2.3.2-2.4.2 hold, and $T \to \infty$.

1. For (2.4.7),

$$\sqrt{T}\left(\widehat{\mathbf{c}} - \mathbf{c}\right) \to_D N\left(\mathbf{0}, \mathbf{\Sigma}^{-1}\mathbf{\Delta}\mathbf{\Sigma}^{-1}\right),$$

   where $\mathbf{\Sigma} = \int_0^1 (\mathbf{\Sigma}_{\mathbf{X}_\mathbf{C}}(\tau) - \mathbf{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau)\mathbf{\Sigma}^{-1}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}(\tau)\mathbf{\Sigma}^{\top}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau))\mathrm{d}\tau$,

$$\mathbf{\Delta} = \int_0^1 \mathbf{P}_\mathbf{C}(\tau)\left(\mathbf{\Sigma}_\mathbf{z}(\tau) \otimes \mathbf{\Omega}(\tau)\right)\mathbf{P}_\mathbf{C}^{\top}(\tau)\mathrm{d}\tau,$$

   $\mathbf{\Sigma}_{\mathbf{X}_\mathbf{C}}(\tau) = \mathbf{C}\mathbf{\Sigma}_\mathbf{Z}(\tau)\mathbf{C}^{\top}$, $\mathbf{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau) = \mathbf{C}\mathbf{\Sigma}_\mathbf{Z}(\tau)\widetilde{\mathbf{C}}^{\top}$, $\mathbf{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}(\tau) = \widetilde{\mathbf{C}}\mathbf{\Sigma}_\mathbf{Z}(\tau)\widetilde{\mathbf{C}}^{\top}$ and $\mathbf{P}_\mathbf{C}(\tau) = \mathbf{C} - \mathbf{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau)\mathbf{\Sigma}^{-1}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}(\tau)\widetilde{\mathbf{C}}$;

2. For (2.4.8),

$$\sqrt{Th}(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) - \frac{1}{2}h^2\tilde{c}_2\boldsymbol{\theta}^{(2)}(\tau) + o_P(h^2)) \to_D N(\mathbf{0}, \tilde{v}_0\mathbf{\Delta}_{\boldsymbol{\theta}}(\tau))$$

   for any given $\tau \in (0,1)$, where $\mathbf{\Delta}_{\boldsymbol{\theta}}(\tau) = \mathbf{\Sigma}^{-1}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}(\tau)\widetilde{\mathbf{C}}\left(\mathbf{\Sigma}_\mathbf{z}(\tau) \otimes \mathbf{\Omega}(\tau)\right)\widetilde{\mathbf{C}}^{\top}\mathbf{\Sigma}^{-1}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}(\tau)$.

Similar to Theorem 2.4.1 (3), both $\mathbf{\Sigma}^{-1}\mathbf{\Delta}\mathbf{\Sigma}^{-1}$ and $\mathbf{\Delta}_{\boldsymbol{\theta}}(\tau)$ can be easily estimated by replacing the unknown quantities with their estimators. In the following section, I conduct numerical studies to evaluate the finite-sample performance of the proposed estimation and inferential methods.

2.5 Numerical Studies

In this section, I first present the details of the numerical implementations in Section 2.5.1, and then conduct extensive simulations in Section 2.5.2. Since the simulation results show that the DWB approach works better numerically than the HAC method, I therefore only apply the DWB approach for the empirical analysis in Section 2.5.3.

2.5.1 Numerical Implementation

I provide some details for practical implementation when applying the results in Section 2.3. The Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u| \leq 1)$ is adopted throughout the numerical studies. For bandwidth selection, since the error innovations, $\mathbf{e}_t = \sum_{j=0}^{\infty} \mathbf{B}_j(\tau_t)\boldsymbol{\varepsilon}_{t-j}$, involved in the time-varying VMA(∞) model are serially correlated, I use the modified cross-validation criterion proposed by Chu and Marron [1991]. Specifically, it is a "leave-(2k+1)-out" version of cross-validation, and $\widehat{h}_{mcv}$ is selected by the following minimization procedure.

$$\widehat{h}_{mcv} = \arg\min_h \sum_{t=1}^{T} \left(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_{k,h}(\tau_t)\right)^\top \left(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_{k,h}(\tau_t)\right), \tag{2.5.1}$$

where $\widehat{\boldsymbol{\mu}}_{k,h}(\tau) = \left[\sum_{t:|t-\tau T|>k} K\left(\frac{\tau_t-\tau}{h}\right)\right]^{-1} \sum_{t:|t-\tau T|>k} \mathbf{x}_t K\left(\frac{\tau_t-\tau}{h}\right)$ and $k=5$.

I then comment on the HAC and DWB methods, of which both require a bandwidth and a kernel function. For the HAC procedure, I use the Bartlett kernel $\psi(x) = (1-|x|)I(|x| \leq 1)$ [e.g., Newey and West, 1987] and the rule of thumb bandwidth $b = 0.75 \cdot (T\widehat{h}_{mcv})^{1/3}$. For the DWB method, I follow the suggestions of Bühlmann [1998], Shao [2010] and Palm et al. [2011] by choosing $\widetilde{h} = c_0 \cdot \widehat{h}_{mcv}^{5/9}$ with $c_0 = 2$, $a(x) = \frac{\int_{-1}^{1} w(u)w(u+|x|)\mathrm{d}u}{\int_{-1}^{1} w^2(u)\mathrm{d}u}$ with $w(u) = \frac{u}{0.43}I\left(u \in [0, 0.43)\right) + I\left(u \in [0.43, 0.57]\right) + \frac{1-u}{0.43}I\left(u \in (0.57, 1]\right)$, and $l = 1.75 \cdot (T\widehat{h}_{mcv})^{1/3}$ respectively.

2.5.2 Simulation Results

I first evaluate the finite sample performance of the DWB and HAC procedures presented in Section 2.3. Consider a multivariate time series with the following data generating process

(DGP)

$$\mathbf{x}_t = \boldsymbol{\mu}(\tau_t) + \mathbf{e}_t, \quad \mathbf{e}_t = \mathbf{A}(\tau_t)\mathbf{e}_{t-1} + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, ..., T, \tag{2.5.2}$$

where $\boldsymbol{\varepsilon}_t$'s are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$, $\boldsymbol{\mu}(\tau) = [\sin(\pi\tau), \cos(\pi\tau)]^\top$, and

$$\mathbf{A}(\tau) = \begin{bmatrix} 0.3\exp(-0.5+\tau) & (\tau-0.5)^3 \\ (\tau-0.5)^3 & 0.3+0.3\sin(\pi\tau) \end{bmatrix}.$$

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

In order to evaluate the finite sample performance, I calculate the point-wise coverage rate associated with $\boldsymbol{\mu}(\cdot)$ based on the HAC estimation method and the DWB procedure with $J = 1000$ bootstrap replications, respectively. Specifically, I consider the coverage at $\tau = 0.1, \ldots, 0.9$, and use the nominal coverage 95%. For each given τ, the coverage probability is first calculated for each component of $\boldsymbol{\mu}(\cdot)$ over 1000 replications, and then I take average across the elements of $\boldsymbol{\mu}(\cdot)$. These probabilities are reported in Table 2.1. It can be seen that the DWB method yields better coverage probabilities, which approach 95% faster than those from the HAC method. For this reason, I will use the DWB method in the empirical study below. In addition, I conjecture that the performance of HAC method can be improved by using a bias corrected trending estimator as explained above. Since the simulation results show that the DWB approach works better than the HAC method numerically, I apply the DWB approach in the empirical analysis.

Table 2.1: Point-wise coverage probabilities for $\boldsymbol{\mu}(\cdot)$

	Mean		Median	
T	DWB	HAC	DWB	HAC
200	0.901	0.818	0.895	0.815
400	0.901	0.833	0.896	0.829
800	0.930	0.868	0.925	0.864

I next evaluate the performance of the semiparametric profile likelihood method for the following DGP:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{B}_1(\tau)x_{t-1} + \boldsymbol{\eta}_t, \tag{2.5.3}$$

Table 2.2: Empirical coverage probabilities for $\mathbf{A}_1$ and $\mathbf{B}_1(\cdot)$

T	$\mathbf{A}_1$	$\mathbf{B}_1(\cdot)$
200	0.952	0.908
400	0.954	0.924
800	0.950	0.939

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

$$\begin{aligned} \mathbf{A}_1 &= \begin{bmatrix} 0.4 & -0.1 \\ -0.1 & 0.4 \end{bmatrix} \\ \mathbf{B}_1(\tau) &= [2\exp(\tau-1)-1, 2\exp(\tau-1)-1]^\top, \end{aligned}$$

in which x_t is an AR(1) process such that $x_t = 0.4x_{t-1} + v_t$ with $v_t \sim N(0,1)$. I set T to be $200, 400, 800$ and conduct 1000 replications for each choice of T. For bandwidth selection, here I use the rule of thumb bandwidth $\widehat{h} = 2.34\sqrt{1/12}T^{-1/5}$ for simplicity.

I evaluate the estimates of $\mathbf{A}_1$ and $\mathbf{B}_1(\tau)$. For each parameter of interest, I report the finite sample coverage probabilities of the confidence intervals. The nominal level is 95%. Specifically, for $\mathbf{B}_1(\tau)$, the coverage probability is first calculated for each functional component over the grid points $\{\tau_t : t = 1, 2, ..., T\}$, and then I further take an average across the elements of $\mathbf{B}_1(\tau)$. After 1000 replications, I present the averaged value of these coverage probabilities in Table 2.2.

As shown in Table 2.2, the finite sample coverage probability of $\mathbf{B}_1(\cdot)$ is smaller than their nominal level when $T = 200$, but are fairly close to 95% as $T = 800$. In addition, the empirical coverage probability of $\mathbf{A}_1$ is very close to the nominal level even when the sample size is relatively small. This result is expected since the rate of convergence on the time-invariant components can be improved to reach a parametric rate.

2.5.3 A Real Data Example

In this subsection, I infer the long-run level of inflation (i.e., trend inflation) and the natural rate of unemployment (NAIRU, which measures the frictional and structural unemployment) based on model (2.3.1). The trend inflation and NAIRU are of central position in setting monetary policy since the Federal Reserve Bank aims to mitigate deviations of inflation and unemployment from their long-run targets (Primiceri, 2006, Stock and Watson, 2016a). The estimation is conducted in exactly the same way as in Section 2.5.1, so I will no longer repeat the details unless necessary.

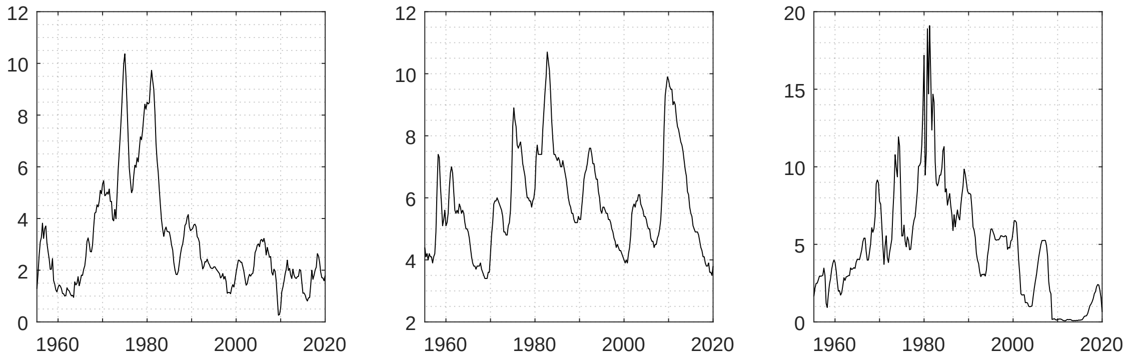

Figure 2.1: Plots of the inflation (left), the unemployment rate (middle) and the interest rate (right)

Specifically, I estimate the time-varying VMA(∞) model (2.3.1) using three commonly adopted macroeconomic variables of the literature (Primiceri, 2005, Cogley et al., 2010), which are the inflation rate (measured by the 100 times the year-over-year log change in the GDP deflator), the unemployment rate and the interest rate (measured by the average value for the Federal funds rates over the quarter). Although I am not interested in the trend of interest rates, I include this variable within the system in order to capture more dynamics and be consistent with the literature. The data are quarterly observations measured at an annual rate from 1954:Q3 to 2020:Q1, which are collected from the Federal Reserve Bank of St. Louis economic database. Figure 2.1 above plots the three variables.

I investigate the trend inflation and the NAIRU. Petrova [2019] considers a Bayesian time-varying VAR(2) model, and induces the long-run mean of $\mathbf{x}_t$ by

$$\boldsymbol{\mu}_t = \lim_{p\to\infty} E_t(\mathbf{x}_{t+p}) = (\mathbf{I}_2 - \mathbf{A}_{1t} - \mathbf{A}_{2t})^{-1}\mathbf{a}_t, \tag{2.5.4}$$

where $\mathbf{a}_t$ is the intercept term, and $\mathbf{A}_{1t}$ and $\mathbf{A}_{2t}$ are the coefficient matrices. The main difference between my method and the Petrova's method is that I explicitly estimate the underlying trends of inflation and unemployment using model (2.3.1).

Figure 2.2 plots the estimates of the trend inflation and the NAIRU (i.e., $\widehat{\boldsymbol{\mu}}(\tau)$), as well as the 95% bootstrap confidence intervals. It is obvious that the underlying trend of inflation is high in the 1970s, but decreases in the subsequent period. After the Great Moderation, the long-run level of inflation is below, but quite close to the Federal Reserve's target of 2%, which indicates that the inflation is more anchored now than that in the 1970s. However, the NAIRU is less persistent and fluctuates over time. In particular, the NAIRU increases rapidly during "Second Oil Crisis" and "Global Financial Crisis".

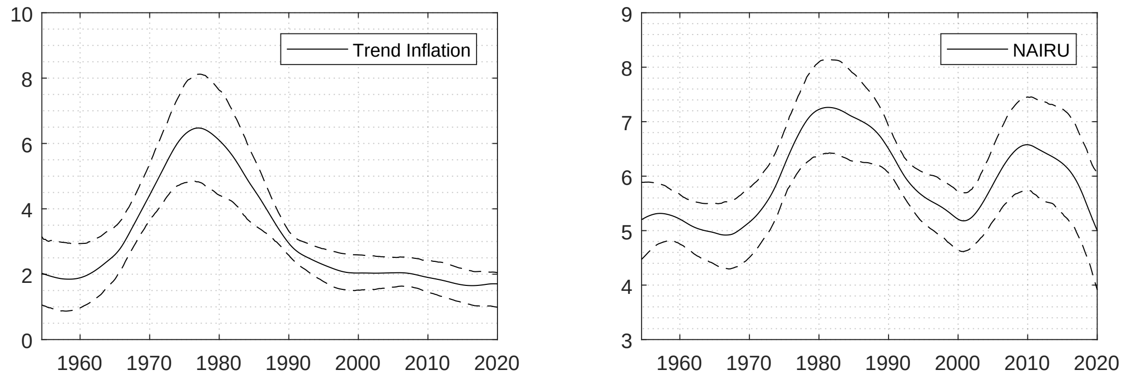

Figure 2.2: The estimated trends (i.e., $\boldsymbol{\mu}(\cdot)$) of inflation and unemployment as well as the associated 95% bootstrap confidence intervals

2.6 Conclusion

In this chapter, I introduce a class of time-varying VMA(∞) processes, and derive a set of asymptotic properties accordingly. My investigation starts with decomposing the weighted sum of time-varying VMA(∞) processes into the long-run and transitory elements, that is known as the BN decomposition [Beveridge and Nelson, 1981, Phillips and Solo, 1992]. As the long-run component of the decomposition yields a martingale approximation, it ensures the feasibility of achieving a variety of asymptotic properties for the multivariate case. Furthermore, I show that these results can be readily applied when establishing inferences for many dynamic time-varying models. In the empirical study, I apply the newly proposed framework to study the long-run level of inflation and the natural rate of unemployment, and find that (1) the long-run level of inflation is more anchored now and is close to the Federal Reserve's target of two percent after the beginning of the Great Moderation period and (2) the natural rate of unemployment is less persistent and increases rapidly during "Second Oil Crisis" and "Global Financial Crisis".

Chapter 3

Estimation, Inference and Empirical Analysis for Time-Varying VAR Models

3.1 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 [2016b] 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 [1996a]. Locally stationary processes are useful when analysing economic and financial time series (Sun et al., 2021a, Xu et al., 2021). For example, the dataset plotted in Figure 3.1 of Section 3.5 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] and Richter and Dahlhaus [2019]. 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 I discuss in this chapter where time-varying parameters are involved in dynamic systems, including VAR models.

In view of the aforementioned issues, this chapter 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 chapter, I specifically study the asymptotic properties of the impulse responses under different identification conditions, which are widely adopted in the literature for different purposes. Also, I 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, I first develop a time-varying vector moving average infinity (VMA(∞)) representation for a class of VAR models. Then I establish uniform consistency and a joint central limit theory for the estimators of VAR coefficients and covariance matrix. Afterwards, I 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, I 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, I 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 chapter, I 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, I use the newly established framework and results to investigate the U.S. government spending multipliers. I 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, my 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 chapter is as follows. Section 3.2 considers a class of time-varying VAR models, and establishes the corresponding asymptotics. Section 3.3 specifically focuses on inferring the time-varying impulse responses subject to different identification schemes. Section 3.4 discusses some implementation issues and presents comprehensive simulations. Section 3.5 provides a case study to demonstrate the empirical relevance. Section 3.6 concludes. Some lengthy mathematical symbols are summarized in Appendix B.1. The preliminary lemmas and the proofs are given in Appendices B.1-B.4.

3.2 The Time-Varying VAR(p) Model

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

$$\mathbf{x}_t = \mathbf{a}(\tau_t) + \sum_{j=1}^{p} \mathbf{A}_j(\tau_t)\mathbf{x}_{t-j} + \boldsymbol{\eta}_t, \tag{3.2.1}$$

where $\tau_t = t/T$, $\boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\varepsilon}_t$, and $\mathbf{a}(\cdot)$, $\mathbf{A}_j(\cdot)$ 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 $\mathbf{x}_t$ has a time-invariant effect on the j^{th} variable of $\mathbf{x}_t$, restricting simultaneous interactions among multiple variables to be time-invariant. For the time being, I assume p is known, and shall come back to its estimation in the end of Section 3.2.1.

The following conditions are necessary for my development.

Assumption 3.2.1.

1. $\mathbf{I}_d - \mathbf{A}_1(\tau)z - \cdots - \mathbf{A}_p(\tau)z^p \neq \mathbf{0}_d$ for all $\tau \in [0,1]$ and all $0 < |z| \leq 1 + \nu$ for some $\nu > 0$. Each element of $\mathbf{A}(\tau) = [\mathbf{a}(\tau), \mathbf{A}_1(\tau), \ldots, \mathbf{A}_p(\tau)]$ is second order continuously differentiable on $[0,1]$ and $\mathbf{A}(\tau) = \mathbf{A}(0)$ 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 3.2.2. $\{\boldsymbol{\varepsilon}_t\}_{t=-\infty}^{\infty}$ is a martingale difference sequence (m.d.s.) adapted to the filtration $\{\mathscr{F}_t\}$, where $\mathscr{F}_t = \sigma(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$ is the σ-field generated by $(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$, $E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top | \mathscr{F}_{t-1}] = \mathbf{I}_d$ almost surely (a.s.), and $\max_t \|\boldsymbol{\varepsilon}_t\|_\delta < \infty$ for some $\delta > 4$.

Assumption 3.2.1.1 infers the local stationarity of (3.2.1). Similar treatments have also been adopted for univariate locally stationary models in the literature (e.g., Assumption

T3 of Zhang and Wu, 2012). Assumption 3.2.1 also allows the underlying data generating process to evolve over time in a smooth manner. The conditions $\mathbf{A}(\tau) = \mathbf{A}(0)$ and $\boldsymbol{\omega}(\tau) = \boldsymbol{\omega}(0)$ for $\tau < 0$ gives

$$\mathbf{x}_t = \mathbf{a}(0) + \sum_{j=1}^{p} \mathbf{A}_j(0)\mathbf{x}_{t-j} + \boldsymbol{\omega}(0)\boldsymbol{\varepsilon}_t, \tag{3.2.2}$$

which basically assumes that $\mathbf{x}_t$ behaves like a parametric VAR(p) model for $t \leq 0$. A similar condition can be found in Vogt [2012] for a nonparametric time series model. Assumption 3.2.2 imposes some conditions on the innovation error terms, which are standard in the VAR literature [Lütkepohl, 2005].

Under these conditions, (3.2.1) admits a time-varying VMA(∞) representation, which sheds light on how to recover the time-varying structural impulse responses. Formally, I present the following proposition.

Proposition 3.2.1. Under Assumptions 3.2.1 and 3.2.2, there exists a time-varying VMA(∞) process of the form:

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

such that $\max_{t\geq 1} \|\mathbf{x}_t - \widetilde{\mathbf{x}}_t\|_\delta = O(T^{-1})$, where $\boldsymbol{\mu}(\tau) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j(\tau)\mathbf{a}(\tau)$, $\boldsymbol{\Psi}_j(\tau) = \mathbf{J}\boldsymbol{\Phi}^j(\tau)\mathbf{J}^\top$, $\mathbf{J} = \left[\mathbf{I}_d, \mathbf{0}_{d\times d(p-1)}\right]$, $\mathbf{B}_j(\tau) = \boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}(\tau)$, and $\boldsymbol{\Phi}(\tau)$ is defined in (B.1.6) for the sake of presentation.

From Proposition 3.2.1, 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 $\mathbf{x}_{t+n}$ is given by $\mathbf{B}_n(\tau_{t+n})\mathbf{e}_j$, where $\mathbf{e}_j$ is a $d \times 1$ selection vector with unity as its j^{th} element and zeros elsewhere. Hence, the impulse responses produced by my 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.

3.2.1 Estimation

To estimate $\{\mathbf{B}_j(\tau) : j \geq 0\}$ via the structural VAR (SVAR) identification schemes, I need a joint central limit theorem for the estimators of the coefficients and the innovation covariance matrix. That said, I consider the estimation of $\mathbf{A}(\cdot)$ and $\boldsymbol{\Omega}(\cdot)$ using the local linear kernel method. Intuitively, when τ_t is in a small neighbmyhood of τ, I can write (3.2.1) as follows:

$$\mathbf{x}_t \approx \left[\mathbf{A}(\tau), h\mathbf{A}^{(1)}(\tau)\right] \mathbf{z}_{t-1}^* + \boldsymbol{\eta}_t, \tag{3.2.3}$$

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

$$\begin{aligned}
\mathrm{vec}[\widehat{\boldsymbol{A}}(\tau)] &= [\mathbf{I}_{d^2p+d}, \mathbf{0}_{d^2p+d}] \cdot \left(\sum_{t=1}^T \mathbf{Z}_{t-1}^* \mathbf{Z}_{t-1}^{*,\top} K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^T \mathbf{Z}_{t-1}^* \mathbf{x}_t K_h(\tau_t - \tau), \\
\widehat{\boldsymbol{\Omega}}(\tau) &= \frac{1}{T}\sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top \omega_t(\tau),
\end{aligned} \tag{3.2.4}$$

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

To facilitate the development, I require the following conditions to hold for the kernel function and the bandwidth.

Assumption 3.2.3. 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 3.2.3 is a set of regular conditions on the kernel function and the bandwidth. With these conditions in hand, I summarize the first theorem of this chapter below.

Theorem 3.2.1. Let Assumptions 3.2.1-3.2.3 hold. Suppose that $\max_{t\geq 1} E[\|\boldsymbol{\varepsilon}_t\|^4 | \mathscr{F}_{t-1}] < \infty$ a.s., and $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$. Then

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

[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 (3.2.4)) is a non-trivial problem, and even has its own literature. I refer interested readers to Zhang and Wu [2012] for more details.

In addition, suppose that conditional on $\mathscr{F}_{t-1}$, the third and fmyth moments of $\boldsymbol{\varepsilon}_t$ are identical to the corresponding unconditional moments a.s., and $Th^5 \to \alpha \in [0, \infty)$. Then for $\forall \tau \in (0, 1)$:

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

where $\mathbf{V}(\tau)$ and $\widehat{\mathbf{V}}(\tau)$ are defined in (B.1.3) and (B.1.5) for the sake of presentation.

The first result of Theorem 3.2.1 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 $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$.

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

$$\widehat{\mathsf{p}} = \underset{1 \le \mathsf{p} \le \mathsf{P}}{\text{argmin}} \left(\log\{\text{RSS}(\mathsf{p})\} + \mathsf{p} \cdot \chi_T\right) \tag{3.2.5}$$

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 (3.2.5).

Theorem 3.2.2. Let Assumptions 3.2.1-3.2.3 hold. Suppose $\max_{t \ge 1} E[\|\boldsymbol{\varepsilon}_t\|^4 | \mathscr{F}_{t-1}] < \infty$ a.s., $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$, $\chi_T \to 0$, and $c_T^{-2}\chi_T \to \infty$, where $c_T = h^2 + \left(\frac{\log T}{Th}\right)^{1/2}$. Then $\Pr\left(\widehat{\mathsf{p}} = p\right) \to 1$.

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, I have estimated all unknown quantities of model (3.2.1).

3.2.2 Inference on Parameter Stability

Before moving on to investigate the impulse responses, I consider a hypothesis test which in practice is able to provide numerical evidence to justify the necessity of the model (3.2.1).

An intuitive question to ask is whether some (if not all) components of the coefficient matrices are time-invariant. Formally, I may write it as follows:

$$\mathbb{H}_0 : \mathbf{C}\boldsymbol{\beta}(\cdot) = \mathbf{c} \text{ for some unknown } \mathbf{c} \in \mathbb{R}^s, \tag{3.2.6}$$

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

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

$$\widehat{Q}_{\mathbf{C},\mathbf{H}} = \int_0^1 \left\{\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\mathbf{c}}\right\}^\top \mathbf{H}(\tau) \left\{\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\mathbf{c}}\right\} \mathrm{d}\tau, \tag{3.2.7}$$

where $\widehat{\boldsymbol{\beta}}(\cdot) := \text{vec}[\widehat{\mathbf{A}}(\cdot)]$ should be obvious, and $\widehat{\mathbf{c}} = \int_0^1 \mathbf{C}\widehat{\boldsymbol{\beta}}(\tau)\mathrm{d}\tau$ is the estimator of $\mathbf{c}$. In (3.2.7), $\mathbf{H}(\cdot)$ is an $s \times s$ positive definite weighting matrix, and is typically set as the precision matrix associated with $\widehat{\boldsymbol{\beta}}(\cdot)$.

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

Theorem 3.2.3. Let Assumptions 3.2.1-3.2.3 hold. Suppose further that $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$, $\max_{t\geq 1} E[\|\boldsymbol{\varepsilon}_t\|^4 | \mathscr{F}_{t-1}] < \infty$ a.s., each element of $\mathbf{A}(\cdot)$ has finite third-order derivative, $Th^2/(\log T)^2 \to \infty$, and $Th^6 \to 0$. If $Th^{5.5} \to 0$ and $E[\|\boldsymbol{\varepsilon}_t\|^\delta | \mathscr{F}_{t-1}] < \infty$ a.s., then I have under $\mathbb{H}_0$

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

where s is the length of $\mathbf{c}$, $\widehat{\mathbf{H}}(\tau) = (\mathbf{C}\widehat{\mathbf{V}}_{\boldsymbol{\beta}}(\tau)\mathbf{C}^\top)^{-1}$, $\widehat{\mathbf{V}}_{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\Sigma}}^{-1}(\tau) \otimes \widehat{\boldsymbol{\Omega}}(\tau)$, $C_B = \int_0^2 (\int_{-1}^{1-v} K(u)K(u+v)\mathrm{d}u)^2 \mathrm{d}v$, and $\widehat{\boldsymbol{\Sigma}}(\tau)$ is defined in (B.1.2).

Theorem 3.2.3 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, I would like to emphasize that the proposed test is in fact a one-side test. Due to the quadratic form of (3.2.7), any departure from the true value will eventually yield a squared term when analysing the asymptotic power. Therefore, the null of (3.2.6) will be rejected at the level α if

$$\widehat{Q}^*_{\mathbf{C},\widehat{\mathbf{H}}} = \frac{T\sqrt{h}\left(\widehat{Q}_{\mathbf{C},\widehat{\mathbf{H}}} - \frac{1}{Th}s\widetilde{v}_0\right)}{\sqrt{4sC_B}} > q_{1-\alpha}, \tag{3.2.8}$$

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

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

$$\mathbb{H}_1 : \mathbf{C}\boldsymbol{\beta}(\tau) = \mathbf{c} + d_T\mathbf{f}(\tau), \tag{3.2.9}$$

where $\mathbf{f}(\tau)$ is a twice continuously differentiable vector of functions, and $d_T \to 0$. The term $d_T\mathbf{f}(\tau)$ characterizes the departure of the time-varying coefficient $\mathbf{C}\boldsymbol{\beta}(\tau)$ from the constant $\mathbf{c}$. Using the development of Theorem 3.2.3, it is straightforward to obtain the following corollary.

Corollary 3.2.1. Let the conditions of Theorem 3.2.3 hold. Under the $\mathbb{H}_1$ of (3.2.9), if $d_T = T^{-1/2}h^{-1/4}$, then

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

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

Corollary 3.2.1 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.,

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

Before I conclude this section, I 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 I have developed and then employed Proposition 3.2.1 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 my proofs may be of general interest and applicability in dealing with similar problems.

3.3 On Impulse Responses

In this section, I consider the estimation and inference of impulse responses. I first study the impulse response using short-run timing and long-run restrictions in Section 3.3.1, 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 I do not repeat them below for the sake of space. In Section 3.3.2, I consider the use of external instruments, which has attracted some attentions recently (e.g., Stock and Watson, 2018, Paul, 2020).

3.3.1 SVAR Identification

As $\mathbf{\Omega}(\cdot) = \boldsymbol{\omega}(\cdot)\boldsymbol{\omega}^\top(\cdot)$, I cannot infer the elements of $\boldsymbol{\omega}(\cdot)$ unless certain identification restrictions are imposed. In the following, I 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 $\widehat{\mathbf{\Omega}}(\tau)$, i.e., $\widehat{\mathbf{\Omega}}(\tau) = \widehat{\boldsymbol{\omega}}(\tau)\widehat{\boldsymbol{\omega}}^\top(\tau)$. Alternatively, one can impose the conditions on the long-run impacts of the shocks (i.e., $\mathbf{B}(\tau)$ defined below). Specifically, define

$$\begin{aligned}
\mathbf{B}(\tau) &:= \sum_{j=0}^{\infty} \mathbf{B}_j(\tau) = \left(\mathbf{I}_d - \sum_{i=1}^{p} \mathbf{A}_i(\tau)\right)^{-1} \boldsymbol{\omega}(\tau), \\
\mathbf{\Psi}(\tau) &:= \sum_{j=0}^{\infty} \mathbf{\Psi}_j(\tau) = \left(\mathbf{I}_d - \sum_{i=1}^{p} \mathbf{A}_i(\tau)\right)^{-1}, \qquad (3.3.1)
\end{aligned}$$

where $\mathbf{B}_j(\cdot)$ and $\mathbf{\Psi}_i(\cdot)$ are defined in Proposition 3.2.1, and the last equalities of both lines follow in an obvious matter. Thus, the elements of $\mathbf{B}(\tau)$ may be recovered from $\mathbf{B}(\tau)\mathbf{B}^\top(\tau) = \mathbf{\Psi}(\tau)\mathbf{\Omega}(\tau)\mathbf{\Psi}^\top(\tau)$. It is then convenient to assume that $\mathbf{B}(\tau)$ is a lower-triangular matrix, so $\widehat{\mathbf{B}}(\tau)$ is obtained from the Cholesky decomposition of $\widehat{\mathbf{\Psi}}(\tau)\widehat{\mathbf{\Omega}}(\tau)\widehat{\mathbf{\Psi}}^\top(\tau)$, where $\widehat{\mathbf{\Psi}}(\tau)$ is defined in the same way as $\mathbf{\Psi}(\tau)$ but replacing $\mathbf{A}_i(\tau)$ with $\widehat{\mathbf{A}}_i(\tau)$. Under the long-run restrictions, $\widehat{\boldsymbol{\omega}}(\tau) = \widehat{\mathbf{\Psi}}^{-1}(\tau)\widehat{\mathbf{B}}(\tau)$.

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

$$\widehat{\boldsymbol{B}}_j(\tau) = \widehat{\mathbf{\Psi}}_j(\tau)\widehat{\boldsymbol{\omega}}(\tau), \tag{3.3.2}$$

where $\widehat{\mathbf{\Psi}}_j(\tau) = \mathbf{J}\widehat{\mathbf{\Phi}}^j(\tau)\mathbf{J}^\top$, $\mathbf{J}$ is defined in Proposition 3.2.1, and $\widehat{\mathbf{\Phi}}(\tau)$ is defined under (B.1.6). Formally, I present the following theorem.

Theorem 3.3.1. Under the conditions of Theorem 3.2.1. For any fixed integer $j \geq 0$

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

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

1. Under the short-run timing restrictions,

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

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

2. Under the long-run restrictions,

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

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

For ease of presentation, I provide the definitions of the respective estimators of $\mathbf{V}(\tau)$ and $\boldsymbol{\Phi}(\tau)$ (i.e., $\widehat{\mathbf{V}}(\tau)$ and $\widehat{\boldsymbol{\Phi}}(\tau)$) in Appendix B.1. 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 $\widehat{\mathbf{V}}(\tau) \to_P \mathbf{V}(\tau)$ by Theorem 3.2.1. As a result, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{B}_j}(\tau) \to_P \boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)$, where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{B}_j}(\tau)$ has a form identical to $\boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)$ but replacing $\boldsymbol{\Phi}(\tau)$, $\boldsymbol{\omega}(\tau)$ and $\mathbf{V}(\tau)$ with their estimators, respectively.

3.3.2 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, I would like to contribute along this line of research by specifically considering the case of Plagborg-Møller and Wolf [2021, p. 971]. Formally, I am interested in the impulse responses of $\mathbf{x}_t$ to the structural shock $\varepsilon_{1,t}$ (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 3.3.1. Suppose π_t can be represented as $\pi_t = \alpha_\pi(\tau_t)\varepsilon_{1,t} + \sum_{j=1}^q \boldsymbol{\beta}_{j,\pi}^\top(\tau_t)\mathbf{x}_{t-j} + e_t$, where $\{e_t\}$ is a sequence of independent variables with $\max_t E|e_t|^\delta < \infty$ and is independent of $\{\boldsymbol{\varepsilon}_t\}$, 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 3.3.1 slightly extends the setting of Plagborg-Møller and Wolf [2021, p. 971] by including the lags of $\mathbf{x}_t$. Using Assumption 3.3.1, simple algebra shows that

$$\frac{E(\eta_{i,t}\pi_t)}{E(\eta_{1,t}\pi_t)} = \frac{\omega_{i,1}(\tau_t)}{\omega_{1,1}(\tau_t)} \quad \text{for } i = 2, \dots, d, \tag{3.3.3}$$

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, I rewrite the model (3.2.1) as

$$\mathbf{x}_t = \boldsymbol{\omega}^*_{\bullet,1}(\tau_t)x_{1,t} + \mathbf{A}^*(\tau_t)\mathbf{z}_{t-1} + \boldsymbol{\eta}^*_t, \tag{3.3.4}$$

where $\boldsymbol{\omega}^*_{\bullet,1}(\tau) = \frac{1}{\omega_{1,1}(\tau)}\boldsymbol{\omega}_{\bullet,1}(\tau)$, $\mathbf{A}^*(\tau) = \mathbf{A}(\tau) - \frac{1}{\omega_{1,1}(\tau)}\boldsymbol{\omega}_{\bullet,1}(\tau)\mathbf{A}_{1,\bullet}(\tau)$, and $\boldsymbol{\eta}^*_t = \sum_{j=2}^d [\boldsymbol{\omega}_{\bullet,j}(\tau_t) - \boldsymbol{\omega}_{\bullet,1}(\tau_t)\frac{\omega_{1,j}(\tau_t)}{\omega_{1,1}(\tau_t)}]\varepsilon_{j,t}$. For model (3.3.4), I immediately obtain that

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

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

$$\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) = [\mathbf{I}_d, \mathbf{0}_d]\left(\mathbf{W}_\tau^\top(\mathbf{I}_{dT} - \mathbf{S})^\top\mathbf{K}_\tau(\mathbf{I}_{dT} - \mathbf{S})\boldsymbol{X}_\tau\right)^{-1}\mathbf{W}_\tau^\top(\mathbf{I}_{dT} - \mathbf{S})^\top\mathbf{K}_\tau(\mathbf{I}_{dT} - \mathbf{S})\mathbf{x}, \tag{3.3.5}$$

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

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

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

With $\widehat{\boldsymbol{\Psi}}_j(\tau)$ defined in (3.3.2) and $\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau)$ in (3.3.5), the SVAR-IV estimator of the impulse response function $\mathbf{B}_{j,\bullet1}(\tau)$ (i.e., the first column of $\mathbf{B}_j(\tau)$) for each $j \geq 0$ is given by

$$\widehat{\boldsymbol{B}}_{j,\boldsymbol{\cdot}1}(\tau) = \widehat{\boldsymbol{\Psi}}_j(\tau)\widehat{\boldsymbol{\omega}}_{\boldsymbol{\cdot},1}(\tau). \tag{3.3.6}$$

I now establish the last theorem of this chapter .

Theorem 3.3.2. Let Assumptions 3.2.1-3.3.1 hold. Suppose that $\max_{t\geq 1} E[\|\boldsymbol{\varepsilon}_t\|^4|\mathscr{F}_{t-1}] < \infty$ 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 $j \geq 0$

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

where

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

For the sake of space, I present the definitions of $\mathbf{V}_{1,1}$ and $\mathbf{V}^*_{2,2}$ in Appendix A. Similar to Theorem 3.3.1, $\boldsymbol{\Sigma}_{\mathbf{B}_j}(\tau)$ can easily be estimated by replacing the unknown quantities with their estimates.

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

3.4 Simulation

In this section, I first provide some details of the numerical implementation in Section 3.4.1, and then respectively examine the estimation and hypothesis testing in Sections 3.4.2 and 3.4.3.

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 (3.2.5), the bandwidth $\widehat{h}_{cv}$ is always chosen by minimizing the following cross-validation criterion function for each p.

$$\widehat{h}_{cv} = \arg\min_h \sum_{t=1}^{T} |\mathbf{x}_t - \widehat{\mathbf{a}}_{-t}(\tau_t) - \sum_{j=1}^{\mathsf{p}} \widehat{\mathbf{A}}_{j,-t}(\tau_t)\mathbf{x}_{t-j}|^2, \tag{3.4.1}$$

where $\widehat{\mathbf{a}}_{-t}(\cdot)$, and $\widehat{\mathbf{A}}_{j,-t}(\cdot)$ are obtained using (3.2.4) but leaving the t^{th} observation out. Once $\widehat{\mathsf{p}}$ and $\widehat{h}_{cv}$ are obtained, the estimation procedure is relatively straightforward.

I now comment on the testing procedure. To improve the finite sample performance of the test, I 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. For simplicity, I adopt the former as follows.

Algorithm - a simulation-assisted testing procedure

Step 1: Use the sample $\{\mathbf{x}_t\}$ to estimate the unrestricted model and the restricted model, and then compute $\widehat{Q}_{\mathbf{C},\widehat{\mathbf{H}}}$ based on (3.2.7).

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

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

Step 4: Repeat Steps 2-3 B times to obtain B bootstrap test statistics $\{\widetilde{Q}^b_{\mathbf{C},\widehat{\mathbf{H}}}\}_{b=1}^B$, as well as its empirical quantile $\widehat{q}_{1-\alpha}$. I reject the null hypothesis (3.2.6) at level α if $\widehat{Q}_{\mathbf{C},\widehat{\mathbf{H}}} > \widehat{q}_{1-\alpha}$.

3.4.2 Examining the Model Estimation

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

$$\mathbf{x}_t = \mathbf{a}(\tau_t) + \mathbf{A}_1(\tau_t)\mathbf{x}_{t-1} + \mathbf{A}_2(\tau_t)\mathbf{x}_{t-2} + \boldsymbol{\eta}_t \quad \text{with} \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\varepsilon}_t \quad \text{for} \quad t = 1, \ldots, T, \tag{3.4.2}$$

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

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

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

First, I check whether the coefficient matrices $\mathbf{A}_1(\tau)$ and $\mathbf{A}_2(\tau)$ satisfy Assumption 3.2.1.1. Thus, for each generated dataset, I 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 3.2.1.1.

Next, I report the percentages of $\widehat{\mathsf{p}} < 2$, $\widehat{\mathsf{p}} = 2$, and $\widehat{\mathsf{p}} > 2$ respectively based on 1000 replications. Table 3.1 shows that the information criterion (3.2.5) performs reasonably well, as the percentages associated with $\widehat{\mathsf{p}} = 2$ are sufficiently close to 1 even for $T = 200$.

Table 3.1: 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, I evaluate the estimates of $\mathbf{A}(\tau)$, $\boldsymbol{\Omega}(\tau)$, as well as the estimates of the impulse responses (say, $\mathbf{B}_1(\tau)$ and $\mathbf{B}_5(\tau)$ without loss of generality) based on the short-run timing restrictions. For each parameter of interest, I 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 $\boldsymbol{\theta}(\cdot) \in \{\mathbf{A}(\cdot), \mathbf{\Omega}(\cdot), \mathbf{B}_1(\tau), \mathbf{B}_5(\tau)\}$, and $\widehat{\boldsymbol{\theta}}^{(n)}(\tau)$ is the estimate of $\boldsymbol{\theta}(\tau)$ for the n^{th} replication. Of interest, I also report the finite sample coverage probabilities of the confidence intervals. The nominal coverage is 95%. Given $\boldsymbol{\theta}(\cdot)$, for each generated dataset, the coverage probability is first calculated for each functional component of $\boldsymbol{\theta}(\cdot)$ over the grid points $\{\tau_t, t = 1, \dots, T\}$, and then I further take an average across the elements of $\boldsymbol{\theta}(\cdot)$. After 1000 replications, I present the averaged value of these coverage probabilities in Table 3.2. As shown in Table 3.2, 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 3.2: The RMSEs and the empirical coverage probabilities (in parentheses)

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

Finally, I evaluate the estimates of the impulse responses $\{\mathbf{B}_{j,\cdot 1}\}$ based on the SVAR-IV method. I generate the instrument variable π_t by $\pi_t = \varepsilon_{1,t} + e_t$, 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 3.2. As shown in Table 3.3, the empirical coverage probabilities are fairly close to 95% when $T = 800$ for all horizons.

Table 3.3: The empirical coverage probabilities of the SVAR-IV estimates of the impulse responses

T	$\mathbf{B}_{1,\cdot 1}(\tau)$	$\mathbf{B}_{3,\cdot 1}(\tau)$	$\mathbf{B}_{5,\cdot 1}(\tau)$	$\mathbf{B}_{7,\cdot 1}(\tau)$
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

3.4.3 Examining the Parameter Stability Test

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

$$\mathbf{x}_t = \mathbf{A}_1(\tau_t)\mathbf{x}_{t-1} + \mathbf{A}_2(\tau_t)\mathbf{x}_{t-2} + \boldsymbol{\eta}_t, \tag{3.4.3}$$

where $\mathbf{A}_2(\cdot)$ and $\boldsymbol{\eta}_t$ are generated in the same way as in Section 3.4.2, and

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

in which $d_T = T^{-1/2}h^{-1/4}$ and b is set to be 0, 2 or 4 in order to investigate the size and local power of the proposed test. I use the proposed testing procedure to test whether the coefficient $\mathbf{A}_1(\cdot)$ is time-varying.

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

$$h = \alpha_1 T^{-1/5}, \quad \alpha_1 = 0.2, 0.4, \ldots, 1.8. \tag{3.4.4}$$

Table 3.4 reports the rejection rates at the 5% and 10% nominal levels. A few facts emerge from the table. First, my test has reasonable sizes using the empirical critical values obtained by the bootstrap procedure. Second, the size behaviour of my 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 my test increases rapidly as b increases.

Table 3.4: 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

3.5 Time-Varying Government Spending Multipliers

In this section, I 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, I address this issue using the newly proposed approach. The estimation is conducted in exactly the same way as in Section 3.4, so I no longer repeat the details.

First, I 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], I 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 3.1 plots the three variables.

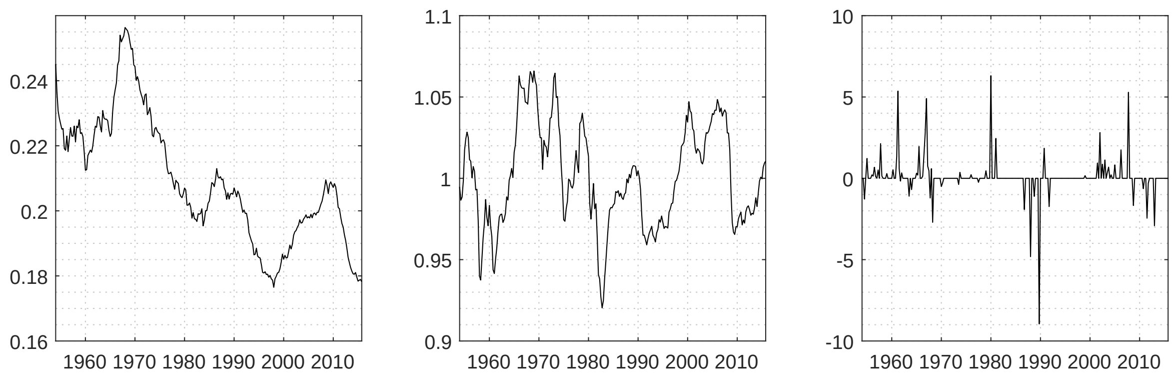

Figure 3.1: Plots of scaled government spending (left), scaled real per capita GDP (middle) and scaled military news (% of GDP, right)

For the time-varying VAR(p) model, the optimal lag is $\widehat{\mathsf{p}} = 2$ by my approach, while it is often assumed to be known with the value varying from 2 to 4 in the literature. Thus,

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

the following analyses focus on the time-varying VAR(2) model (referred to as TV-VAR(2) hereafter). I 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 $|\lambda_{\max}\{\widehat{\boldsymbol{\Phi}}(\tau)\}|$ varies from 0.7-0.9 over time, and thus Assumption 3.2.1.1 is automatically met. In addition, I use the 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, I further check whether the VAR coefficients (i.e., the policy transmission mechanism) are time-varying. I employ the proposed test statistic to examine the constancy of model coefficients, and summarize the results in Table 3.5. From Table 3.5 (the row labelled "Constancy"), I choose the TV-VAR(2) model over a constant VAR model. I then apply the testing procedure to distinguish time-varying intercept or time-varying autoregressive coefficients. By Table 3.5, at the 5% significance level I 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 3.5: Summary statistics of the test

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

[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 my main goal, I no longer conduct more testing along this line.

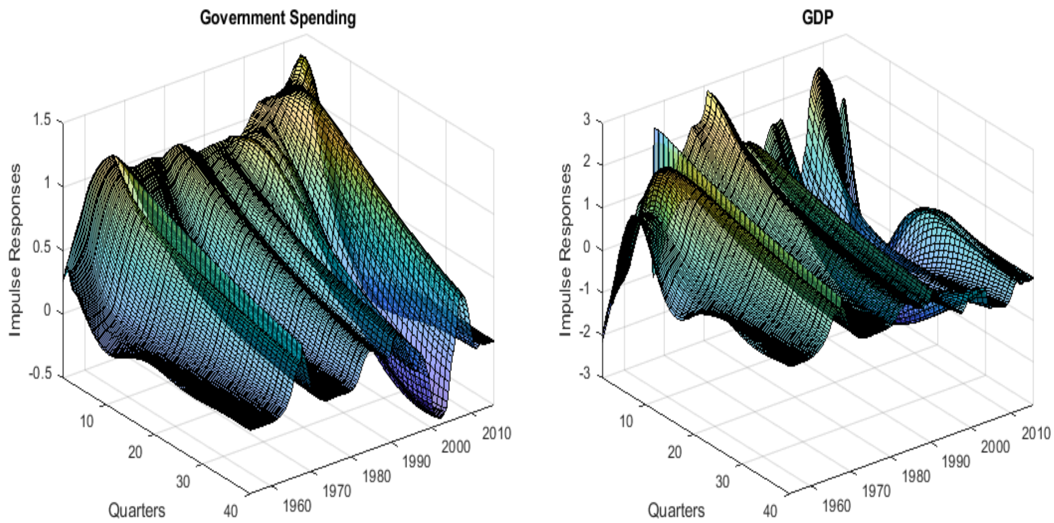

Figure 3.2: Time-varying impulse responses to a military news shock. Left front axis left: Quarters (Horizon). Right front axis left: Time.

I now discuss how government spending and GDP respond to a military news shock over time. Figure 3.2 plots the time-varying impulse responses of government spending and 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, my 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.

I 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 3.3 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 3.3.2. From Figure 3.3, 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 reduc-*

ing multipliers in recent years." Importantly, my results provide a solid evidence for this argument.

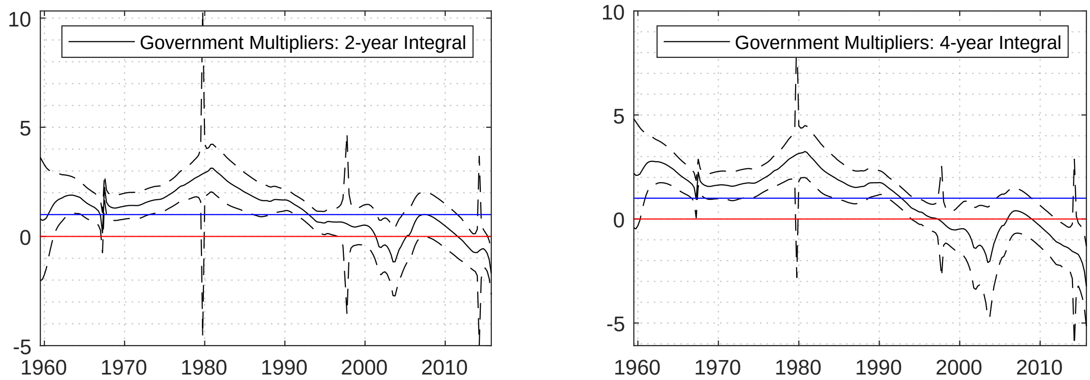

Figure 3.3: Estimates of time-varying government multipliers based on SVAR-IV approach for a 2-year and 4-year horizon, as well as their 95% condidence intervals.

Finally, I explore the sensitivity of my findings to different identification conditions. Alternatively, I 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 3.4 plots the estimates of time-varying multipliers based on the short-run timing restrictions. Figure 3.4 reveals similar results that government multipliers decreasing over time and are not significantly different from zeros after 1990s.

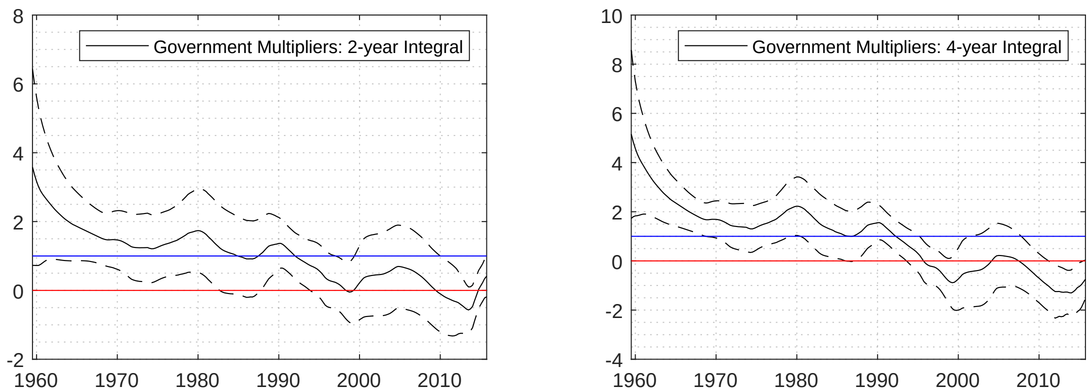

Figure 3.4: Estimates of time-varying government multipliers based on short-run timing restriction for a 2-year and 4-year horizon, as well as their 95% condidence intervals.

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

3.6 Conclusion

In this chapter, I investigate 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, I establish 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, I demonstrate the empirical relevance of the proposed methods through an application to estimating time-varying government multipliers using U.S. data, and find that the government spending multipliers are above one before 1990s and are not significantly different from zero after 1990s using different identification schemes.

Chapter 4

Time-Varying Multivariate Causal Processes

4.1 Introduction

The family of vector autoregressive (VAR) models and the family of multivariate (G)ARCH models are among some of the most popular frameworks for modelling dynamic interactions of multiple variables. The VAR family usually captures the dynamic by imposing structures on the time series itself, while the (G)ARCH family imposes restrictions on the conditional second moments. I acknowledge the vast literature of both families, and have no intention to exhaust all relevant studies in this chapter for the sake of space. I refer interested readers to Stock and Watson [2001] and Bauwens et al. [2006] for excellent review on both families.

Although both families have rich literature on their own, to the best of the authors' knowledge not many works have been done to bridge them. Among limited attempts (e.g., Ling and McAleer, 2003, Bardet and Wintenberger, 2009), most (if not all) of these studies rely on the stationarity assumption. While the stationarity assumption comes in handy when deriving asymptotic properties, it may not be very realistic in practice (e.g., Preuss et al., 2015, Chen et al., forthcoming . For example, economic and financial data always include different macro shocks, as a consequence the behaviour can be quite volatile; the climate data may contain certain time trend which recently has attracted lots of attention due to green house emission; etc. Either way, certain nonstationarity may always occur.

To account for the nonstationarity, locally stationary processes have received considerable attention since the seminal work of Dahlhaus [1996b], Dette et al. [2011], Zhang and Wu [2012], Truquet [2017], Dahlhaus et al. [2019], among others. In contrast to the unit root process, the locally stationary process nicely balances the stationarity and nonstationarity by allowing for the simultaneous presence of both types of behaviours in one time series process. In a very recent paper, Karmakar et al. [forthcoming] consider simultaneous inference for a general class of univariate p-Markov processes with time-varying coefficients, which covers several time-varying versions of the classical univariate models (e.g., AR, ARCH, AR-ARCH) as special cases. Despite its great generality, their study still rules out the time-varying versions of some widely used models (e.g., ARMA, GARCH, ARMA-GARCH). Also, it is worth mentioning this line of research heavily focuses on univariate time series, which somewhat limits the popularity of locally stationary processes.

That said, it is reasonable to call for a framework which can marry the VAR family and the (G)ARCH family while allowing for nonstationarity. To provide a concrete example, consider a time-varying multivariate GARCH model, which can model the co-movements of financial returns. Detailed investigation on such a model can help answer research questions like (i). Is the volatility of a market leading the volatility of other markets? (ii) Whether the correlations between asset returns change over time? (iii). Are they increasing in the long run, perhaps because of the globalization of financial markets? These are of great practical importance for both investors and policymakers (Bauwens et al., 2006, Diebold and Yilmaz, 2009).

To allow for flexibility as much as possible from the modelling perspective, I consider a class of multivariate causal processes as follows:

$$\mathbf{x}_t = \begin{cases} \boldsymbol{\mu}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \boldsymbol{\theta}(\tau_t)\right) + \mathbf{H}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \boldsymbol{\theta}(\tau_t)\right) \boldsymbol{\varepsilon}_t, & \text{for} \quad t = 1, \ldots, T, \\ \boldsymbol{\mu}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \boldsymbol{\theta}(0)\right) + \mathbf{H}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \boldsymbol{\theta}(0)\right) \boldsymbol{\varepsilon}_t & \text{for} \quad t \leq 0, \end{cases} \tag{4.1.1}$$

where $\tau_t = t/T$, $\boldsymbol{\mu}(\cdot)$ is an m-dimensional random vector, $\mathbf{H}(\cdot)$ is an $m \times m$-dimensional random matrix, $\boldsymbol{\theta}(\tau)$ is a $d \times 1$ time-varying parameter of interest with each element belonging to $C^3[0,1]$, and $\{\boldsymbol{\varepsilon}_t\}$ is a sequences of independent and identically distributed (i.i.d.) error innovations. Note that the value of d usually depends on the value of m, and the connection becomes clear once a specific model is considered. As far as I am concerned, both of m and d are fixed throughout the chapter. Notably, both $\boldsymbol{\mu}(\cdot)$ and $\mathbf{H}(\cdot)$ are known, and share the same unknown parameter $\boldsymbol{\theta}(\cdot)$. The setting for $t \leq 0$ regulates the time series for the periods that I do not observe, which is commonly adopted when certain nonstationarity gets involved (e.g., Vogt, 2012). Essentially, it requires the initial time period does not have a diverging behaviour.

Before proceeding further, I provide two examples to briefly illustrate the rationality behind (4.1.1), and leave the detailed investigation on these examples to Section 4.2.4. I refer interested readers to Ling [2003], Ling and McAleer [2003] and Bardet and Wintenberger [2009] for extensive investigation on the parametric counterparts of these examples.

Example 1: Consider the time-varying VARMA(p, q) model

$$\mathbf{x}_t = \mathbf{a}(\tau_t) + \sum_{j=1}^{p} \mathbf{A}_j(\tau_t)\mathbf{x}_{t-j} + \boldsymbol{\eta}_t + \sum_{j=1}^{q} \mathbf{B}_j(\tau_t)\boldsymbol{\eta}_{t-j} \quad \text{with} \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\varepsilon}_t. \tag{4.1.2}$$

It is not hard to show that (4.1.2) admits a presentation in the form of (4.1.1), and

$$\boldsymbol{\theta}(\tau) = \text{vec}(\mathbf{a}(\tau), \mathbf{A}_1(\tau), \ldots, \mathbf{A}_p(\tau), \mathbf{B}_1(\tau), \ldots, \mathbf{B}_q(\tau), \boldsymbol{\Omega}(\tau)), \tag{4.1.3}$$

where $\boldsymbol{\Omega}(\cdot) := \boldsymbol{\omega}(\cdot)\boldsymbol{\omega}^\top(\cdot)$.

Example 2: Consider the time-varying multivariate GARCH(p, q) model

$$\begin{aligned} \mathbf{x}_t &= \text{diag}(h_{1,t}^{1/2}, \ldots, h_{m,t}^{1/2})\boldsymbol{\eta}_t, \\ \mathbf{h}_t &= \mathbf{c}_0(\tau_t) + \sum_{j=1}^{p} \mathbf{C}_j(\tau_t)\left(\mathbf{x}_{t-j} \odot \mathbf{x}_{t-j}\right) + \sum_{j=1}^{q} \mathbf{D}_j(\tau_t)\mathbf{h}_{t-j}, \end{aligned} \tag{4.1.4}$$

where $h_{j,t}$ stands for the j^{th} element of $\mathbf{h}_t$, and $\boldsymbol{\eta}_t = \boldsymbol{\Omega}^{1/2}(\tau_t)\boldsymbol{\varepsilon}_t$. The model (4.1.4) generalizes the models of Bollerslev [1990] and Jeantheau [1998]. Similar to Example 1, I show that (4.1.4) admits a representation in the form of (4.1.1), and

$$\boldsymbol{\theta}(\tau) = \text{vec}(\mathbf{c}_0(\tau), \mathbf{C}_1(\tau), \ldots, \mathbf{C}_p(\tau), \mathbf{D}_1(\tau), \ldots, \mathbf{D}_q(\tau), \boldsymbol{\Omega}(\tau)). \tag{4.1.5}$$

In view of the development of Example 1 and Example 2 in Section 4.2.4, one may further show the time-varying counterparts of the parametric models mentioned in Bardet and Wintenberger [2009] are also covered by (4.1.1). To this end, I argue that (4.1.1) does not only allows for nonstationarity and conditional heteroskedasticity, but also provides sufficient flexibility to cover many well adopted models in the literature.

In this chapter, my contributions are in the following four-fold: (1). I consider a wide class of time-varying multivariate causal processes which nests many classic and new examples as special cases; (2). I prove the existence of a weakly dependent stationary approximation for the model (4.1.1) at any given time of interest (i.e., $\forall \tau \in [0,1]$), which is the foundation in order to establish asymptotic properties associated with the model; (3). I establish the estimation theory, and provide both point-wise and simultaneous inferences on the coefficient functions of which both are important for practical works (Zhou and Wu, 2010); (4). I demonstrate the theoretical findings through both simulated and real data examples.

The chapter is organized as follows. Section 4.2 presents the theoretical findings associated with the stationary approximation, estimation and inferences. In Section 4.3, I conduct extensive simulation studies to examine the theoretical findings, and further investigate the time-varying volatility spillover effects and conditional correlations between the Chinese and U.S. Stock market. Section 4.4 concludes. Some mathematical tools and the proofs of main results are summarized in Appendix C.1-C.2. Some preliminary lemmas and their proofs are given in Appendix C.3-C.4.

4.2 Estimation and Asymptotics

In this section, I first prove the existence of a weakly dependent stationary approximation for the model (4.1.1) in Section 4.2.1; I then provide the estimation approach using the local linear quasi-maximum-likelihood estimation and establish the asymptotic properties of the proposed estimator in Section 4.2.2; Section 4.2.3 provides results on both point-wise and simultaneous inferences; Section 4.2.4 gives some detailed examples to justify the usefulness of my study.

4.2.1 Stationary Approximation

To study (4.1.1), the first challenge lies in the fact that the model may not be stationary. Therefore, for $\forall \tau \in [0,1]$, I initialize my analysis by finding a stationary approximation for each $\mathbf{x}_t$ with $t \geq 1$. By doing so, I are able to measure the weak dependence of $\{\mathbf{x}_t\}$ using the nonlinear system introduced in Wu [2005], which then provides us a framework to derive the asymptotic properties accordingly.

To be clear on the dependence measure, consider an example in which $\mathbf{e}_t$ is a stationary process, and admits a causal representation $\mathbf{e}_t = \mathbf{J}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$ with $\mathbf{J}(\cdot)$ being a measurable function. For $k \geq 0$, I define the following dependence measure:

$$\delta_r^{\mathbf{e}}(k) = \|\mathbf{J}(\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_{k-1}, \ldots) - \mathbf{J}(\boldsymbol{\varepsilon}_k, \ldots, \boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_0^*, \boldsymbol{\varepsilon}_{-1}, \ldots)\|_r\,, \tag{4.2.1}$$

where $\boldsymbol{\varepsilon}_0^*$ is an independent copy of $\boldsymbol{\varepsilon}_0$. Being able to measure the time series dependence such as (4.2.1) is the starting point for time series analyses.

That said, I now introduce some basic assumptions.

Assumption 4.2.1.

1. $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of i.i.d. random variables with $E(\boldsymbol{\varepsilon}_1) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^\top) = \mathbf{I}_m$, and $\|\boldsymbol{\varepsilon}_1\|_r < \infty$ for some $r \geq 2$.

2. For $\forall \mathbf{z}, \mathbf{z}' \in (\mathbb{R}^m)^\infty$ and $\forall \boldsymbol{\vartheta} \in \mathbb{R}^d$, there exist nonnegative sequences $\{\alpha_j(\boldsymbol{\vartheta})\}_{j=1}^\infty$ and $\{\beta_j(\boldsymbol{\vartheta})\}_{j=1}^\infty$ such that

$$|\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}) - \boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta})| \leq \sum_{j=1}^\infty \alpha_j(\boldsymbol{\vartheta})|\mathbf{z}_j - \mathbf{z}_j'|,$$
$$|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) - \mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})| \leq \sum_{j=1}^\infty \beta_j(\boldsymbol{\vartheta})|\mathbf{z}_j - \mathbf{z}_j'|,$$

   where $\mathbf{z}_j$ and $\mathbf{z}_j'$ are the j^{th} columns of $\mathbf{z}$ and $\mathbf{z}'$ respectively.

3. For $\forall \tau \in [0,1]$, $\boldsymbol{\theta}(\tau)$ lies in the interior of $\boldsymbol{\Theta}_r$, where

$$\boldsymbol{\Theta}_r := \left\{ \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \mid \sum_{j=1}^\infty \alpha_j(\boldsymbol{\vartheta}) + \|\boldsymbol{\varepsilon}_1\|_r \sum_{j=1}^\infty \beta_j(\boldsymbol{\vartheta}) < 1 \right\}$$

and $\boldsymbol{\Theta}$ is a compact set of $\mathbb{R}^d$.

Assumption 4.2.1.1 is standard when studying dynamic time series model (Lütkepohl, 2005).

In Assumption 4.2.1.2, $\boldsymbol{\vartheta}$ is a generic $d \times 1$ vector, and has the same length as $\boldsymbol{\theta}(\cdot)$. This assumption imposes Lipschitz-type conditions on $\boldsymbol{\mu}(\cdot)$ and $\mathbf{H}(\cdot)$, which are rather minor, and can be easily fulfilled by a variety of models such as those mentioned in Section 4.1. See Propositions 4.2.3-4.2.4 below for details.

Assumption 4.2.1.3 does not only guarantee a stationary approximation for each $\mathbf{x}_t$, but also ensures the approximated process has some proper moments. Similar conditions have also been adopted in Bardet and Wintenberger [2009].

With these conditions in hand, I present the following proposition which facilitates the development in what follows.

Proposition 4.2.1. Let Assumption 4.2.1 hold. For any $\tau \in [0,1]$, there exists a stationary process

$$\widetilde{\mathbf{x}}_t(\tau) = \boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \ldots; \boldsymbol{\theta}(\tau)\right) + \mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \ldots; \boldsymbol{\theta}(\tau)\right) \boldsymbol{\varepsilon}_t$$

such that

1. $\sup_{\tau\in[0,1]} \|\widetilde{\mathbf{x}}_t(\tau)\|_r < \infty$,

2. $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k) \leq O(1) \inf_{1\leq p\leq k}\{\rho(\tau)^{k/p} + \sum_{j=p+1}^{\infty} [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))]\} \to 0$ as $k \to \infty$,

where $\rho(\tau) := \sum_{j=1}^{\infty} \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \sum_{j=1}^{\infty} \beta_j(\boldsymbol{\theta}(\tau))$.

It is worth mentioning that for a univariate p-Markov process

$$\widetilde{x}_{p,t}(\tau) = \mu\left(\widetilde{x}_{t-1}(\tau), \ldots, \widetilde{x}_{t-p}(\tau); \boldsymbol{\theta}(\tau)\right) + H\left(\widetilde{x}_{t-1}(\tau), \ldots, \widetilde{x}_{t-p}(\tau); \boldsymbol{\theta}(\tau)\right) \varepsilon_t,$$

Karmakar et al. [forthcoming] show that there exists $0 < \rho < 1$ such that $\sup_{\tau\in[0,1]} \delta_r^{\widetilde{x}_p(\tau)}(k) = O(\rho^k)$ based on the development of Wu and Shao [2004]. From a methodological viewpoint, I give a set of new proofs which allow us to measure the dependence of multivariate causal processes with infinity memory. The term $\sum_{j=p+1}^{\infty}[\alpha_j(\boldsymbol{\theta}(\tau))+\beta_j(\boldsymbol{\theta}(\tau))]$ in the second result of Proposition 4.2.1 arises due to the infinity memory structure of $\widetilde{\mathbf{x}}_t(\tau)$. Thus, the dependence $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k)$ relies on the choice of p and the decay rates of the coefficients $\alpha_j(\boldsymbol{\theta}(\tau))$ and $\beta_j(\boldsymbol{\theta}(\tau))$.

To ensure $\widetilde{\mathbf{x}}_t(\tau)$ can approximate $\mathbf{x}_t$ reasonably well, I impose more structure below.

Assumption 4.2.2.

1. There exists a nonnegative sequence $\{\chi_j\}$ with $\sum_{j=1}^{\infty}\chi_j < \infty$ such that for $\forall \mathbf{z} \in (\mathbb{R}^m)^\infty$ and $\forall \boldsymbol{\vartheta}, \boldsymbol{\vartheta}' \in \boldsymbol{\Theta}_r$

$$|\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}')|+|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}')| \leq |\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|\sum_{j=1}^{\infty}\chi_j|\mathbf{z}_j|.$$

2. Let $\sup_{\tau\in[0,1]}\alpha_j(\boldsymbol{\theta}(\tau)) = O(j^{-(2+s)})$ and $\sup_{\tau\in[0,1]}\beta_j(\boldsymbol{\theta}(\tau)) = O(j^{-(2+s)})$ for some $s > 0$.

Assumption 4.2.2.1 imposes another Lipschitz-type condition with respect to the parameter space. Assumption 4.2.2.2 further restricts the decay rates of $\alpha_j(\boldsymbol{\theta}(\tau))$ and $\beta_j(\boldsymbol{\theta}(\tau))$.

Using Assumptions 4.2.1-4.2.2, I can measure the distance between $\widetilde{\mathbf{x}}_t(\tau)$ and $\mathbf{x}_t$ as follows.

Proposition 4.2.2. Suppose Assumptions 4.2.1-4.2.2 hold. Then

1. $\|\widetilde{\mathbf{x}}_1(\tau)-\widetilde{\mathbf{x}}_1(\tau')\|_r = O(|\tau-\tau'|)$ for $\forall \tau, \tau' \in [0,1]$,

2. $\max_{t\geq 1}\|\mathbf{x}_t-\widetilde{\mathbf{x}}_t(\tau_t)\|_r = O(T^{-1})$.

I can consider Proposition 4.2.2 as the stochastic version of the Hölder continuity. Having established the stationary approximation in Proposition 4.2.2, I move on to investigate the estimation theory in the next subsection.

4.2.2 Estimation

I point out a few facts to facilitate the setup of the likelihood function. First, let $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots)$ include all the information of $\mathbf{x}_t$ up to the time period t. However, in practice, my observation on $\mathbf{x}_t$ only starting from $t = 1$, so I have to work with the truncated version of $\mathbf{z}_t$ for each $t \geq 1$:

$$\mathbf{z}_t^c = (\mathbf{x}_t, \ldots, \mathbf{x}_1, \mathbf{0}, \ldots).$$

Second, I note that when τ_t is sufficiently close to τ,

$$\boldsymbol{\theta}(\tau_t) \approx \boldsymbol{\theta}(\tau) + h\boldsymbol{\theta}^{(1)}(\tau) \cdot \frac{\tau_t - \tau}{h}.$$

Therefore, I am able to parametrize $\boldsymbol{\theta}(\cdot)$, and consider the maximum-likelihood estimation for each given τ. Finally, since $\boldsymbol{\varepsilon}_t$ may not be normally distributed, I consider the local linear quasi-maximum-likelihood estimation (QMLE) method.

Thus, the likelihood function is specified as follows:

$$\mathscr{L}_\tau(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2) = \frac{1}{T}\sum_{t=1}^{T} \ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) K_h(\tau_t - \tau), \tag{4.2.2}$$

where

$$\begin{aligned} \ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \boldsymbol{\vartheta}) \quad = \quad & -\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu}(\mathbf{z}_{t-1}^c; \boldsymbol{\vartheta}))^\top \left(\mathbf{H}(\mathbf{z}_{t-1}^c; \boldsymbol{\vartheta})\mathbf{H}(\mathbf{z}_{t-1}^c; \boldsymbol{\vartheta})^\top\right)^{-1} (\mathbf{x}_t - \boldsymbol{\mu}(\mathbf{z}_{t-1}^c; \boldsymbol{\vartheta})) \\ & -\frac{1}{2} \log\det \left(\mathbf{H}(\mathbf{z}_{t-1}^c; \boldsymbol{\vartheta})\mathbf{H}(\mathbf{z}_{t-1}^c; \boldsymbol{\vartheta})^\top\right). \end{aligned}$$

Accordingly, for $\forall \tau$, $(\boldsymbol{\theta}(\tau), h\boldsymbol{\theta}^{(1)}(\tau))$ is estimated by

$$(\widehat{\boldsymbol{\theta}}(\tau), \widehat{\boldsymbol{\theta}}^\star(\tau)) = \underset{(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2) \in \mathbf{E}_T(r)}{\operatorname{argmax}} \mathscr{L}_\tau(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2), \tag{4.2.3}$$

where $\mathbf{E}_T(r) = \boldsymbol{\Theta}_r \times (h \cdot \boldsymbol{\Theta}^{(1)})$ and $\boldsymbol{\Theta}^{(1)}$ is a compact set.

I impose more structures in order to derive the asymptotic distribution.

Assumption 4.2.3.

1. $\inf_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_r, \mathbf{z} \in (\mathbb{R}^m)^\infty} \lambda_{\min}\left(\mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta})\mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta})^\top\right) \geq \underline{c}$ for some $\underline{c} > 0$.

2. For any $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_r$, $\boldsymbol{\mu}(\widetilde{\mathbf{z}}_t(\tau); \boldsymbol{\theta}(\tau)) = \boldsymbol{\mu}(\widetilde{\mathbf{z}}_t(\tau); \boldsymbol{\vartheta})$ and $\mathbf{H}(\widetilde{\mathbf{z}}_t(\tau); \boldsymbol{\theta}(\tau)) = \mathbf{H}(\widetilde{\mathbf{z}}_t(\tau); \boldsymbol{\vartheta})$ a.s. imply $\boldsymbol{\vartheta} = \boldsymbol{\theta}(\tau)$ for some t, where $\widetilde{\mathbf{z}}_t(\tau) = (\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{x}}_{t-1}(\tau), \ldots)$.

Assumption 4.2.4.

1. $\boldsymbol{\mu}(\cdot;\boldsymbol{\vartheta})$ and $\mathbf{H}(\cdot;\boldsymbol{\vartheta})$ are twice continuously differentiable with respect to $\boldsymbol{\vartheta}$.

2. There exists a nonnegative sequence $\{\chi_j\}_{j=1}^{\infty}$ with $\chi_j = O(j^{-(2+s)})$ and some $s > 0$ such that for any $\mathbf{z}, \mathbf{z}' \in (\mathbb{R}^m)^\infty$ and any $\boldsymbol{\vartheta}, \boldsymbol{\vartheta}' \in \boldsymbol{\Theta}_r$:

$$|\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}) - \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}')| + |\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) - \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}')| \leq |\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'| \sum_{j=1}^{\infty} \chi_j |\mathbf{z}_j|,$$

$$|\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}) - \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta})| + |\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) - \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^k \mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})| \leq \sum_{j=1}^{\infty} \chi_j |\mathbf{z}_j - \mathbf{z}_j'|,$$

where $\boldsymbol{\nabla}_{\boldsymbol{\vartheta}} = \left(\frac{\partial}{\partial \vartheta_1}, \ldots, \frac{\partial}{\partial \vartheta_d}\right)^\top$, and $k = 1, 2$.

Assumption 4.2.5. 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 4.2.3.1 ensures the positive definiteness of the covariance matrix of the likelihood function, and is widely adopted when studying the multivariate time series (e.g., page 2736 of Bardet and Wintenberger, 2009). In fact, the validity of this assumption is easy to justify in view of (4.2.7) and (4.2.9) for Example 1 and Example 2 below. Assumption 4.2.3.2 imposes an standard identification condition in the literature of M-estimation (e.g., Proposition 3.4 of Jeantheau, 1998). It is noteworthy that the current form of Assumption 4.2.3 accommodates the flexibility of the model (4.1.1), which is in fact unnecessary if I have a detailed model in practice. See Section 4.2.4 for example.

Assumption 4.2.4 imposes the Lipschitz-type conditions on the first and second order derivatives of $\boldsymbol{\mu}(\cdot)$ and $\mathbf{H}(\cdot)$ to ensure the smoothness of their functional components.

Assumption 4.2.5 is a set of regular conditions on the kernel function and the bandwidth.

With these conditions in hand, I summarize the first theorem of this chapter below.

Theorem 4.2.1. Suppose Assumptions 4.2.1-4.2.5 hold with $r \geq 6$.

(1). If $Th^7 \to 0$, then for any $\tau \in (0, 1)$

$$\sqrt{Th}\left(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) - \frac{1}{2}h^2 \widetilde{c}_2 \boldsymbol{\theta}^{(2)}(\tau)\right) \to_D N\left(\mathbf{0}, \widetilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\tau)\right),$$

where $\mathbf{\Sigma}_{\boldsymbol{\theta}}(\tau) = \mathbf{\Sigma}^{-1}(\tau)\mathbf{\Omega}(\tau)\mathbf{\Sigma}^{-1}(\tau)$, $\mathbf{\Sigma}(\tau) = E\left(\nabla^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau); \boldsymbol{\theta}(\tau))\right)$ and

$$\mathbf{\Omega}(\tau) = E\left(\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau); \boldsymbol{\theta}(\tau)) \cdot \nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau); \boldsymbol{\theta}(\tau))^\top\right).$$

(2). In addition, if $\boldsymbol{\varepsilon}_t$ is normally distributed, I have $\mathbf{\Omega}(\tau) = -\mathbf{\Sigma}(\tau)$ and thus $\mathbf{\Sigma}_{\boldsymbol{\theta}}(\tau) = \mathbf{\Omega}^{-1}(\tau)$.

After deriving the asymptotic distribution, I establish both the point-wise inference and the simultaneous inference in the following.

4.2.3 Inference

In this section, I first discuss how to conduct point-wise inference, and then move on to derive the asymptotic results associated with the simultaneous inference. Notably, the simultaneous inference nests the traditional constancy test as a special case. It does not only allow one to examine whether a time-varying model should be preferred to its parametric counterpart, but also allows one to test any particular functional form of interest. Zhou and Wu [2010] discuss the importance of simultaneous inference in length, so I refer interested readers to their paper for more informative explanation.

Point-wise Inference: First, I construct a bias-corrected estimator in order to remove the asymptotic bias of Theorem 4.2.1. Specifically, I let

$$\widetilde{\boldsymbol{\theta}}(\tau) = 2\widehat{\boldsymbol{\theta}}_{h/\sqrt{2}}(\tau) - \widehat{\boldsymbol{\theta}}(\tau), \tag{4.2.4}$$

where $\widehat{\boldsymbol{\theta}}_{h/\sqrt{2}}(\tau)$ is defined in the same way as $\widehat{\boldsymbol{\theta}}(\tau)$ but using the bandwidth $h/\sqrt{2}$.

After tedious development (Lemma A.1.8 of Appendix B), I have uniformly over $\tau \in [h, 1-h]$

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) &= -\mathbf{\Sigma}^{-1}(\tau)\frac{1}{Th}\sum_{t=1}^{T}\widetilde{K}((\tau_t-\tau)/h)\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \\ &\quad + O_P((Th)^{-1/2}h^{3/2}(\log T)^{1/2}) + o(h^3),\end{aligned}$$

where $\widetilde{K}(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x)$ that is essentially a fourth-order kernel. It then infers that under the conditions of Theorem 4.2.1,

$$\sqrt{Th}(\widetilde{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)) \to_D N\left(\mathbf{0}, \widetilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\tau)\right).$$

It is noteworthy that the construction of (4.2.4) is different from directly using the fourth-order kernel in the regression. In terms of bandwidth selection, the traditional methods (e.g., cross-validation) still remain valid for (4.2.4) (Richter and Dahlhaus, 2019). However, if one directly employs the fourth-order kernel in the regression, it remains unclear how to select the optimal bandwidth in practice.

Now I discuss how to estimate $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\tau)$ which is constructed by $\boldsymbol{\Sigma}(\tau)$ and $\boldsymbol{\Omega}(\tau)$. Intuitively, I consider the following estimator

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}(\tau) = \widehat{\boldsymbol{\Sigma}}^{-1}(\tau)\widehat{\boldsymbol{\Omega}}(\tau)\widehat{\boldsymbol{\Sigma}}^{-1}(\tau), \tag{4.2.5}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}(\tau) &= A_T(\tau)^{-1} \sum_{t=1}^{T} \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^2 \ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \widehat{\boldsymbol{\theta}}(\tau)) K_h(\tau_t - \tau), \\
\widehat{\boldsymbol{\Omega}}(\tau) &= A_T(\tau)^{-1} \sum_{t=1}^{T} \boldsymbol{\nabla}_{\boldsymbol{\vartheta}} \ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \widehat{\boldsymbol{\theta}}(\tau)) \cdot \boldsymbol{\nabla}_{\boldsymbol{\vartheta}} \ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \widehat{\boldsymbol{\theta}}(\tau))^\top K_h(\tau_t - \tau), \\
A_T(\tau) &= \sum_{t=1}^{T} K_h(\tau_t - \tau).
\end{aligned}$$

Note that I consider a local constant estimator in (4.2.5) rather than a local linear one, that is to avoid an implementation issue for finite sample studies (i.e., nonpositive definite covariance may occur when the local linear approach is employed). Such a numerical problem has been well explained and investigated in the literature. See Chen and Leng [2015] for example.

The following corollary summarizes the asymptotic property of (4.2.5).

Corollary 4.2.1. Let Assumptions 4.2.1-4.2.5 hold with $r \geq 6$. Suppose further that

$$\sup_{\tau \in [0,1]} [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(5/2+s)})$$

for some $s > 0$. In addition, let $h(\log T)^2 \to 0$ and $T^{1-6/r}h \to \infty$. Then

$$\sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}(\tau) - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\tau)| = o_P(1).$$

Simultaneous Inference: I now consider the simultaneous inference. To allow for flexibility, I first introduce a selection matrix $\mathbf{C}$ with full row rank, which selects the parameters of interest as follows:

$$\boldsymbol{\theta}_{\mathbf{C}}(\tau) := \mathbf{C}\boldsymbol{\theta}(\tau).$$

Accordingly, the estimator and the corresponding asymptotic covariance matrix become

$$\widehat{\boldsymbol{\theta}}_{\mathbf{C}}(\tau) := \mathbf{C}\widehat{\boldsymbol{\theta}}(\tau) \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathbf{C}}(\tau) = \mathbf{C}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\tau)\mathbf{C}^\top.$$

Theorem 4.2.2. Let Assumptions 4.2.1-4.2.5 hold with $r \geq 6$. Suppose further that

$$\sup_{\tau\in[0,1]} [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(3+s)})$$

for some $s > 0$. In addition, let $(\log T)^4/(T^\nu h) \to 0$ with $\nu = \frac{1}{2} - \frac{r-6}{4rs/3+2r-4}$ and $Th^7 \log T \to 0$. Then

$$\lim_{T\to\infty} \Pr\left(\sqrt{\frac{Th}{\widetilde{v}_0}} \sup_{\tau\in[h,1-h]} \left|\boldsymbol{\Sigma}_{\mathbf{C}}^{-1/2}(\tau)\left\{\widehat{\boldsymbol{\theta}}_{\mathbf{C}}(\tau) - \boldsymbol{\theta}_{\mathbf{C}}(\tau) - \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\theta}_{\mathbf{C}}^{(2)}(\tau)\right\}\right| \right.$$
$$\left. -B(1/h) \leq \frac{u}{\sqrt{2\log(1/h)}}\right) = \exp(-2\exp(-u)),$$

where $C_K = \frac{\{\int_{-1}^{1} |K^{(1)}(u)|^2 \mathrm{d}u/\widetilde{v}_0\pi\}^{1/2}}{\Gamma(k/2)}$,

$$B(1/h) = \sqrt{2\log(1/h)} + \frac{\log(C_K) + (k/2 - 1/2)\log(\log(1/h)) - \log(2)}{\sqrt{2\log(1/h)}}$$

and $\Gamma(\cdot)$ is the Gamma function.

In Theorem 4.2.2, ν is slightly smaller than $1/2$ as I only require r to be slightly larger than 6. Hence, the usual optimal bandwidth $h_{opt} = O(T^{-1/5})$ satisfies the conditions $(\log T)^4/(T^\nu h) \to 0$ and $Th^7 \log T \to 0$.

As shown in Theorem 4.2.2, the convergence rate of the simultaneous confidence intervals for $\boldsymbol{\theta}_{\mathbf{C}}(\cdot)$ is of logarithmic rate and is therefore slow. In order to improve the rate, I consider a bootstrap method which shows much better finite sample performance. I summarize the result in the following corollary.

Corollary 4.2.2. Under the conditions of Theorem 4.2.2. Suppose that $h = O(T^{-\kappa})$ with $1/7 < \kappa < \nu$. Then, on a richer probability space, there exists i.i.d. k-dimensional standard normal variables $\mathbf{v}_1, \ldots, \mathbf{v}_T$ such that

$$\sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\theta}}_{\mathbf{C}}(\tau) - \boldsymbol{\theta}_{\mathbf{C}}(\tau) - \frac{1}{2}h^2 b_h(\tau)\boldsymbol{\theta}_{\mathbf{C}}^{(2)}(\tau) - \boldsymbol{\Sigma}_{\mathbf{C}}^{1/2}(\tau)\mathbf{V}_h^*(\tau)| = O_P\left(\frac{T^{-\alpha}}{\sqrt{Th}\log T}\right),$$

where $\alpha = \min\{(\nu-\kappa)/2, (7\kappa-1)/2, \kappa/2\}$, $\widetilde{c}_{k,h}(\tau) = \int_{-\tau/h}^{(1-\tau)/h} u^k K(u)\mathrm{d}u$,

$$b_h(\tau) = \frac{\widetilde{c}_{2,h}^2(\tau) - \widetilde{c}_{1,h}(\tau)\widetilde{c}_{3,h}(\tau)}{\widetilde{c}_{0,h}(\tau)\widetilde{c}_{2,h}(\tau) - \widetilde{c}_{1,h}^2(\tau)},$$

$\mathbf{V}_h^*(\tau) = T^{-1}\sum_{t=1}^T \mathbf{v}_t\omega_{t,h}(\tau)$, and $\omega_{t,h}(\tau) = K_h(\tau_t - \tau)\frac{\widetilde{c}_{2,h}(\tau) - \frac{\tau_t-\tau}{h}\widetilde{c}_{1,h}(\tau)}{\widetilde{c}_{0,h}(\tau)\widetilde{c}_{2,h}(\tau) - \widetilde{c}_{1,h}^2(\tau)}$ is the local linear weight.

By Corollary 4.2.2, I propose the following numerical procedure to construct the simultaneous confidence band (SCB) of $\boldsymbol{\theta}_{\mathbf{C}}(\tau)$:

Step 1 Use the sample $\{\mathbf{x}_t\}_{t=1}^T$ to estimate $\widehat{\boldsymbol{\theta}}_{\mathbf{C}}(\tau)$ by (4.2.3), and compute $\widetilde{\boldsymbol{\theta}}_{\mathbf{C}}(\tau)$ based on (4.2.4).

Step 2 Generate i.i.d. k-dimensional standard normal variables $\{\mathbf{v}_t^*\}$ and calculate the quantity $\sup_{\tau\in[0,1]} |\mathbf{V}_h^*(\tau)|$, in which $\mathbf{V}_h^*(\tau) = T^{-1}\sum_{t=1}^T \mathbf{v}_t^*(2\omega_{t,h/\sqrt{2}}(\tau) - \omega_{t,h}(\tau))$.

Step 3 Repeat Step 2 R times to obtain the empirical $(1-\alpha)^{th}$ quantile $\widehat{q}_{1-\alpha}$ of $\sup_{\tau\in[0,1]} |\mathbf{V}_h^*(\tau)|$.

Step 4 Calculate $\widehat{\boldsymbol{\Sigma}}_{\mathbf{C}}(\tau)$ using (4.2.5) and construct the SCB of $\boldsymbol{\theta}_{\mathbf{C}}(\tau)$ by $\widetilde{\boldsymbol{\theta}}_{\mathbf{C}}(\tau) + \widehat{\boldsymbol{\Sigma}}_{\mathbf{C}}^{1/2}(\tau)\widehat{q}_{1-\alpha}\mathbb{B}_k$, where $\mathbb{B}_k = \{\mathbf{u} \in \mathbb{R}^k : |\mathbf{u}| \le 1\}$ is the unit ball, and k is the rank of $\mathbf{C}$.

4.2.4 Examples

Below, I demonstrate the usefulness of the aforementioned results by considering Example 1 and Example 2 of Section 4.1.

Example 1 (Cont.) - For $\forall\tau \in [0,1]$, simple algebra shows that the approximated stationary process is defined by

$$\widetilde{\boldsymbol{x}}_t(\tau) = \boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \ldots; \boldsymbol{\theta}(\tau)\right) + \mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \ldots; \boldsymbol{\theta}(\tau)\right)\boldsymbol{\varepsilon}_t, \tag{4.2.6}$$

where $\boldsymbol{\theta}(\tau)$ has been defined in (4.1.1), and

$$\begin{aligned} \boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \ldots; \boldsymbol{\theta}(\tau)\right) &= \mathbf{B}_\tau^{-1}(1)\mathbf{a}(\tau) + \sum_{j=1}^{\infty} \boldsymbol{\Gamma}_j(\tau)\widetilde{\boldsymbol{x}}_{t-j}(\tau), \\ \mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \ldots; \boldsymbol{\theta}(\tau)\right) &= \boldsymbol{\omega}(\tau). \end{aligned} \tag{4.2.7}$$

Additionally, in (4.2.7), $\boldsymbol{\Gamma}_j(\tau)$ is yielded as follows:

$$\mathbf{I}_m - \sum_{j=1}^{\infty} \boldsymbol{\Gamma}_j(\tau) L^j = \mathbf{B}_\tau^{-1}(L)\mathbf{A}_\tau(L),$$

where $\mathbf{A}_\tau(L) := \mathbf{I}_m - \mathbf{A}_1(\tau)L - \cdots - \mathbf{A}_p(\tau)L^p$ and $\mathbf{B}_\tau(L) := \mathbf{I}_m + \mathbf{B}_1(\tau)L + \cdots + \mathbf{B}_q(\tau)L^q$.

Then I am able to present the following proposition.

Proposition 4.2.3. Let $\|\boldsymbol{\varepsilon}_t\|_r < \infty$ for some $r > 4$. Suppose that there is a compact set

$$\boldsymbol{\Theta} = \{\boldsymbol{\vartheta} = \text{vec}(\mathbf{a}, \mathbf{A}_1, \ldots, \mathbf{A}_p, \mathbf{B}_1, \ldots, \mathbf{B}_q, \boldsymbol{\Omega}) \mid \boldsymbol{\vartheta} \in \mathbb{R}^d\}$$

such that (1). for $\forall\tau \in [0,1]$, $\boldsymbol{\theta}(\tau)$ lies in the interior of $\boldsymbol{\Theta}$, (2). $\det(\mathbf{A}(L)\mathbf{B}(L)) \neq 0$ for all $|L| \le 1$, (3). $\boldsymbol{\Omega} > 0$, where $\mathbf{A}(L) := \mathbf{I}_m - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p$ and $\mathbf{B}(L) := \mathbf{I}_m + \mathbf{B}_1 L + \cdots + \mathbf{B}_q L^q$ are coprime and satisfy some necessary identification conditions. Then, the results of Theorems 4.2.1 and 4.2.2 hold for model (4.1.2).

I note that the detailed identification conditions required for VARMA processes (e.g., the final equations form or echelon form) can be found in Lütkepohl [2005]. I no longer discuss them here in order not to deviate from my main goal.

Example 2 (Cont.) - I further let

$$\mathbf{\Omega}(\tau)=\begin{bmatrix}1 & \rho_{1,2}(\tau) & \cdots & \rho_{1,m}(\tau)\\ \rho_{1,2}(\tau) & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \rho_{m-1,m}(\tau)\\ \rho_{1,m}(\tau) & \rho_{m-1,m}(\tau) & \ddots & 1\end{bmatrix}.$$

For $\forall\tau\in[0,1]$, the corresponding approximated stationary process is defined as

$$\widetilde{\boldsymbol{x}}_t(\tau)=\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\ldots;\boldsymbol{\theta}(\tau)\right)\boldsymbol{\varepsilon}_t, \tag{4.2.8}$$

where

$$\begin{aligned}&\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\ldots;\boldsymbol{\theta}(\tau)\right)\\ =\ &\mathrm{diag}^{1/2}\left(\mathbf{D}_\tau^{-1}(1)\mathbf{c}_0(\tau)+\sum_{j=1}^{\infty}\mathbf{\Psi}_j(\tau)\left(\widetilde{\mathbf{x}}_{t-j}(\tau)\odot\widetilde{\mathbf{x}}_{t-j}(\tau)\right)\right).\end{aligned} \tag{4.2.9}$$

Note that $\mathbf{\Psi}_j(\tau)$ is generated as follows:

$$\mathbf{\Psi}_\tau(L):=\mathbf{I}_m-\sum_{j=1}^{\infty}\mathbf{\Psi}_j(\tau)L^j=\mathbf{D}_\tau^{-1}(L)\mathbf{C}_\tau(L),$$

where $\mathbf{C}_\tau(L):=\mathbf{C}_1(\tau)L+\cdots+\mathbf{C}_p(\tau)L^p$ and $\mathbf{D}_\tau(L):=\mathbf{I}_m-\mathbf{D}_1(\tau)L-\cdots-\mathbf{D}_q(\tau)L^q$.

Consequently, I can present the following proposition.

Proposition 4.2.4. Suppose that there is a compact set

$$\mathbf{\Theta}=\{\boldsymbol{\vartheta}=\mathrm{vec}(\mathbf{c}_0,\mathbf{C}_1,\ldots,\mathbf{C}_p,\mathbf{D}_1,\ldots,\mathbf{D}_q,\mathbf{\Omega})\mid\boldsymbol{\vartheta}\in\mathbb{R}^d\}$$

such that (1). for $\forall\tau\in[0,1]$, $\boldsymbol{\theta}(\tau)$ lies in the interior of $\mathbf{\Theta}$, (2). $\|\mathbf{\Omega}^{1/2}\boldsymbol{\varepsilon}_t\|_r^2\sum_{j=1}^{\infty}|\mathbf{\Psi}_j|<1$ for some $r>6$, (3). all the roots of $|\mathbf{I}_m-\sum_{j=1}^{p}\mathbf{C}_j-\sum_{j=1}^{q}\mathbf{D}_j|$ are outside the unit circle with $\mathbf{C}_j$'s and $\mathbf{D}_j$'s being nonnegative matrices, (4). $\mathbf{c}_0$ is a vector of positive elements, (5). $\mathbf{C}(L)$ and $\mathbf{D}(L)$ are coprime and the formulation of the GARCH part is minimal, where $\mathbf{C}(L):=\mathbf{C}_1L+\cdots+\mathbf{C}_pL^p$ and $\mathbf{D}(L):=\mathbf{I}_m-\mathbf{D}_1L-\cdots-\mathbf{D}_qL^q$. Then the results at

Theorems 4.2.1 and 4.2.2 hold for model (4.1.4).

For the identification conditions of the GARCH process, I refer readers to Proposition 3.4 of Jeantheau [1998], who proves that assuming the minimal representation is enough for ensuring Assumption 4.2.3 holds.

In the following section, I conduct numerical studies using both simulated and real data to evaluate the finite-sample performance of the proposed estimation and inferential methods.

4.3 Numerical Studies

In this section, I first present the details of the numerical implementations in Section 4.3.1, and then conduct extensive simulations in Section 4.3.2. Section 4.3.3 presents a real data example on the conditional correlations between the Chinese and U.S. stock markets.

4.3.1 Numerical Implementation

Throughout the numerical studies, the Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u| \leq 1)$ is adopted. Following Zhou and Wu [2010], I use $\widetilde{h} = 2\widehat{h}$ for the biased corrected estimator, where $\widehat{h}$ is the bandwidth selected by the cross-validation method of Richter and Dahlhaus [2019].

Specifically, define the leave-one-out local linear QMLE

$$(\widehat{\boldsymbol{\theta}}_{h,-t}(\tau), h\widehat{\boldsymbol{\theta}}^{(1)}_{h,-t}(\tau)) = \underset{(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)\in\mathbf{E}_T(r)}{\operatorname{argmax}} \mathscr{L}^c_{T,-t}(\tau,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2), \tag{4.3.1}$$

where $\mathscr{L}^c_{T,-t}(\tau,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2) = \frac{1}{T}\sum_{s=1,\neq t}^{T} \ell(\mathbf{x}_s, \mathbf{z}^c_{s-1}; \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_s - \tau)/h) K_h(\tau_s - \tau)$. Then, the bandwidth is chosen by

$$\widehat{h} = \underset{h}{\operatorname{argmax}}\, T^{-1} \sum_{t=1}^{T} \ell(\mathbf{x}_t, \mathbf{z}^c_{t-1}; \widehat{\boldsymbol{\theta}}_{h,-t}(\tau_t)). \tag{4.3.2}$$

As shown in Richter and Dahlhaus [2019], this cross validation method works well as long as $\boldsymbol{\nabla}\ell$ is uncorrelated, which implies that this desirable property should hold in my case.

Notably, when considering some specific models, the implementation may be further simplified. I provide more discussions along this line in Appendix B.4.

4.3.2 Simulation Results

In the simulation studies, I examine the empirical coverage probabilities of simultaneous confidence intervals for nominal levels $\alpha = 90\%,\ 95\%$. I consider the time-varying VARMA$(2,1)$ and multivariate GARCH$(1,1)$ model as follows:

1. DGP 1 : $\mathbf{x}_t = a_1(\tau_t)\mathbf{x}_{t-1} + a_2(\tau_t)\mathbf{x}_{t-2} + \boldsymbol{\eta}_t + \mathbf{B}_1(\tau_t)\boldsymbol{\eta}_{t-1}, \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\varepsilon}_t$, where $\{\boldsymbol{\varepsilon}_t\}$ are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$, $a_1(\tau) = 0.6\exp(\tau - 1)$, $a_2(\tau) = -0.3\exp(\tau - 1)$,

$$\begin{aligned}\mathbf{B}_1(\tau) &= \begin{bmatrix} 0.5\exp\{\tau - 0.5\} & -0.8(\tau - 0.5)^2 \\ -0.8(\tau - 0.5)^2 & 0.5 + 0.3\sin(\pi\tau) \end{bmatrix}, \\ \boldsymbol{\omega}(\tau) &= \begin{bmatrix} 1.5 + 0.2\exp\{0.5 - \tau\} & 0 \\ 0.2\exp\{0.5 - \tau\} & 1.5 + 0.5(\tau - 0.5)^2 \end{bmatrix}.\end{aligned}$$

   Here I use final equations form to ensure the uniqueness of the VARMA representation.

2. DGP 2 : $\mathbf{x}_t = \mathrm{diag}(h_{1,t}^{1/2}, \ldots, h_{m,t}^{1/2})\boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t = \boldsymbol{\Omega}^{1/2}(\tau_t)\boldsymbol{\varepsilon}_t$, $\mathbf{h}_t = \mathbf{c}_0(\tau_t) + \mathbf{C}_1(\tau_t)\left(\mathbf{x}_{t-1} \odot \mathbf{x}_{t-1}\right) + \mathbf{D}_1(\tau_t)\mathbf{h}_{t-1}$, $\{\boldsymbol{\varepsilon}_t\}$ are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$, $\mathbf{c}_0(\tau) = [2\exp\{0.5\tau - 0.5\}, 3 + 0.2\cos(\tau)]^\top$,

$$\begin{aligned}\mathbf{C}_1(\tau) &= \begin{bmatrix} 0.4 + 0.05\cos(\tau) & 0.05(\tau - 0.5)^2 \\ 0.05(\tau - 0.5)^2 & 0.4 + 0.05\sin(\tau) \end{bmatrix}, \\ \mathbf{D}_1(\tau) &= \begin{bmatrix} 0.4 - 0.1\cos(\tau) & 0 \\ 0 & 0.3 - 0.1\sin(\tau) \end{bmatrix}, \\ \boldsymbol{\Omega}(\tau) &= \begin{bmatrix} 1 & 0.3\sin(\tau) \\ 0.3\sin(\tau) & 1 \end{bmatrix}.\end{aligned}$$

Let the sample size be $T \in \{500, 1000\}$ ($T \in \{1000, 2000, 4000\}$) for the VARMA model (the GARCH model). I conduct 1000 replications for each choice of T. Several different bandwidths close to $\widetilde{h}$ are reported to check the sensitivity of bandwidth selection.

I present the empirical coverage probabilities associated with the SCB in Tables 4.1-4.2. For the vector- or matrix-valued unknown coefficients, I take an average across the elements. A few facts emerge from the tables. First, the finite sample coverage probabilities are smaller than their nominal level when $T = 500$ ($T = 1000, 2000$) for the VARMA model (the GARCH model), but are fairly close to their nominal level as $T = 1000$ ($T = 4000$) for the VARMA model (the GARCH model). Second, the behaviour of the estimated simultaneous confidence intervals is not sensitive to the choices of bandwidths. Third, the GARCH model requires more data to reach a reasonable finite sample performance.

Table 4.1: Empirical Coverage Probabilities of the SCB for DGP 1

		90%				95%			
	$\widetilde{h}$	$\alpha_1(\cdot)$	$\alpha_2(\cdot)$	$\mathbf{B}_1(\cdot)$	$\mathbf{\Omega}(\cdot)$	$\alpha_1(\cdot)$	$\alpha_2(\cdot)$	$\mathbf{B}_1(\cdot)$	$\mathbf{\Omega}(\cdot)$
$T = 500$	0.35	0.845	0.877	0.821	0.847	0.905	0.915	0.889	0.905
	0.4	0.865	0.875	0.847	0.876	0.912	0.930	0.897	0.909
	0.45	0.862	0.895	0.847	0.878	0.915	0.930	0.898	0.919
	0.5	0.875	0.895	0.847	0.876	0.905	0.945	0.901	0.920
$T = 1000$	0.3	0.895	0.925	0.887	0.884	0.960	0.960	0.947	0.947
	0.35	0.910	0.927	0.886	0.890	0.940	0.967	0.940	0.930
	0.4	0.917	0.939	0.901	0.899	0.947	0.959	0.948	0.939
	0.45	0.937	0.932	0.908	0.895	0.957	0.957	0.947	0.937

Table 4.2: Empirical Coverage Probabilities of the SCB for DGP 2

		90%				95%			
	$\widetilde{h}$	$\mathbf{c}_0(\cdot)$	$\mathbf{C}_1(\cdot)$	$\mathbf{D}_1(\cdot)$	$\mathbf{\Omega}(\cdot)$	$\mathbf{c}_0(\cdot)$	$\mathbf{C}_1(\cdot)$	$\mathbf{D}_1(\cdot)$	$\mathbf{\Omega}(\cdot)$
$T = 1000$	0.55	0.802	0.810	0.784	0.889	0.869	0.876	0.838	0.945
	0.60	0.824	0.820	0.791	0.879	0.882	0.866	0.843	0.945
	0.65	0.832	0.820	0.796	0.874	0.889	0.872	0.859	0.945
	0.70	0.820	0.823	0.792	0.879	0.892	0.881	0.871	0.950
$T = 2000$	0.50	0.827	0.835	0.841	0.889	0.897	0.881	0.901	0.950
	0.55	0.829	0.825	0.843	0.884	0.892	0.881	0.903	0.940
	0.60	0.849	0.833	0.871	0.900	0.900	0.888	0.910	0.950
	0.65	0.852	0.835	0.873	0.910	0.907	0.889	0.910	0.950
$T = 4000$	0.35	0.879	0.879	0.882	0.869	0.929	0.932	0.943	0.920
	0.4	0.899	0.879	0.882	0.859	0.950	0.944	0.943	0.919
	0.45	0.904	0.899	0.879	0.838	0.950	0.947	0.946	0.950
	0.50	0.867	0.857	0.884	0.898	0.929	0.944	0.946	0.960

4.3.3 A Real Data Example

In this subsection, I investigate the time-varying conditional correlations between the Chinese and U.S. stock markets using the time-varying multivariate GARCH model. Recently, there is a growing literature to invest the relationship of the two stock markets (e.g., Zhang and Li, 2014, Pan et al., 2022), as the Chinese stock market has become the world's second largest stock market after 2009. Understanding the interactions among different financial markets is important for investors and policymakers [Diebold and Yilmaz, 2009, BenSaïda, 2019]. For example, high equity market interdependence implies poor diversification benefits from portfolios, but highlights the possibility of better hedging benefits.

Previous research documents a strong positive link between the degree of globalization and equity market interdependence [Baele, 2005]. Along this line of research, one important question is that whether the interdependence between the Chinese and U.S. stock markets has increased over time due to globalization so that estimates from historical data are unreliable for modern policy analysis, asset pricing and risk management. The existing results present many discrepancies, which may be due to the fact that the relationship evolves with time. In what follows, I address this issue using the newly proposed approach. The estimation is conducted in exactly the same way as in Section 4.3.1, so I no longer repeat the details.

I calculate the Chinese and U.S. stock returns based on weekly Shanghai Stock Exchange (SSE) Composite Index and S&P 500 Index as they are the most comprehensive and diversified stock indices. The sample employed in this study spanning from January 2000 to February 2022 provides 1119 observations[1]. Figure 4.1 plots the two weekly returns as well as sample autocorrelation functions of squared data, which shows the typical "volatility clustering" phenomenon.

[1] The data are collected from Yahoo Finance at `https://finance.yahoo.com/`.

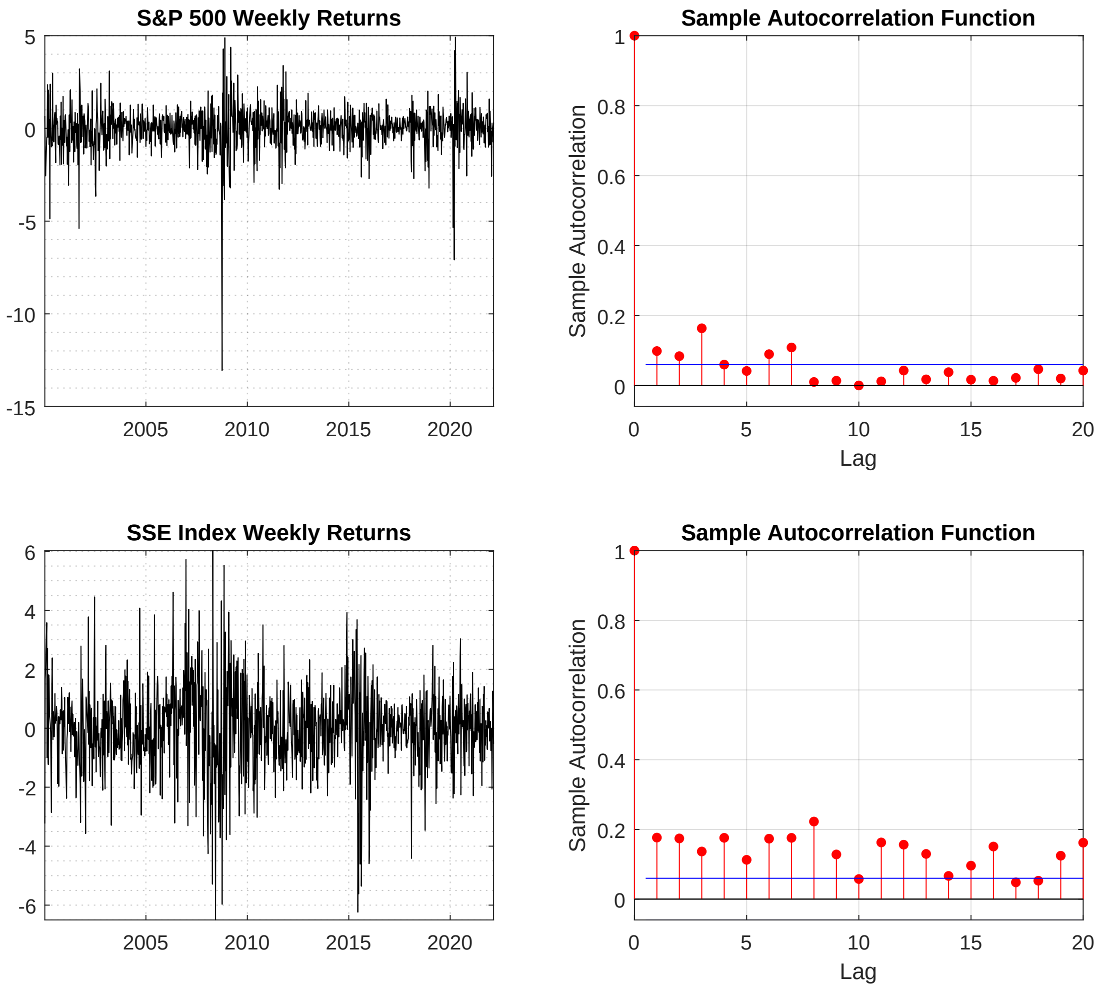

Figure 4.1: S&P 500 and SSE Index returns as well as sample autocorrelation functions of squared data

I next fit the data to a time-varying multivariate GARCH(1,1) model. Figure 4.2 plots the estimates (black solid line) of time-varying conditional correlations between the two stock markets as well as 95% simultaneous confidence intervals (red dashed line) and 95% pointwise confidence intervals (black dashed line). Based on the simultaneous confidence intervals, apparently, the conditional correlations vary with respect to time. Moreover, as clearly presented in Figure 4.2, the interdependence between the two stock markets is increasing over time. By examining the pointwise confidence intervals, I can conclude that the two stock markets are not significantly correlated before 2005, but the relationship has been greatly enhanced in recent years. These results have important implications for investment and risk management. For example, it implies that the Chinese and U.S. investors who use cross-country portfolio strategies to eliminate country specific risks may be benefit from hedging. However, all types of investors should be cautious since the relations between the Chinese and U.S. stock markets are time-varying.

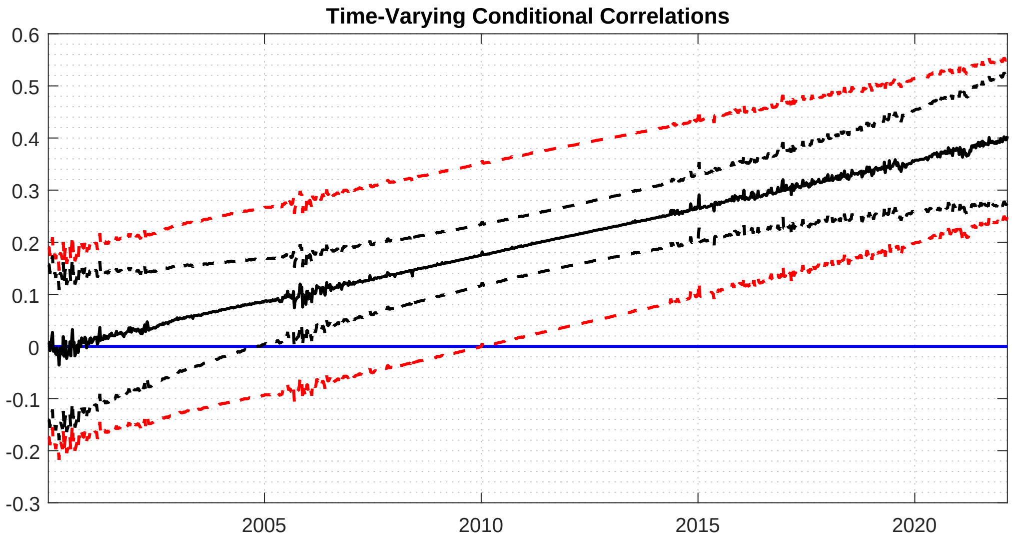

Figure 4.2: Time-varying conditional correlations between the Chinese and U.S. stock markets

4.4 Conclusion

In this chapter, I consider a wide class of time-varying multivariate causal processes which nests many classic and new examples as special cases. I first prove the existence of a weakly dependent stationary approximation for the model (4.1.1) which is the foundation to establish the corresponding asymptotic properties. Afterwards, I consider the QMLE estimation approach, and provide both point-wise and simultaneous inferences on the coefficient functions. In addition, I demonstrate the theoretical findings through both simulated and real data examples. In particular, I show the empirical relevance of my study using an application to evaluate the conditional correlations between the stock markets of China and U.S. I find that the interdependence between the two stock markets is increasing over time.

Chapter 5

Conclusion and Future Work

This dissertation contributes to the estimation and inferential theory of multivariate time series models with local stationarity. The second chapter has proposed a new class of time-varying VMA(∞) processes. The third chapter has introduced a new class of time-varying VAR models in which the VAR coefficients and covariance matrix of the error innovations are allowed to change smoothly over time. The fourth chapter has considered a wide class of time-varying multivariate causal processes which nests many classic and new examples as special cases. The asymptotic properties of the proposed estimators and test statistics are established in each chapter. Monte Carlo studies are conducted to check the finite sample performance of the proposed tests and estimators. Empirical applications are conducted in each chapter to illustrate the usefulness of the proposed models and methods.

There are of course many possible extensions. For example, 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; the third one is to consider quantile regression methods for such locally stationary multivariate causal processes; etc. I wish to leave these challenging issues for future study in my academic career.

Bibliography

Lieven Baele. Volatility spillover effects in european equity markets. *Journal of Financial and Quantitative Analysis*, 40(2):373–401, 2005.

Jean-Marc Bardet and Olivier Wintenberger. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Annals of Statistics*, 37(5B): 2730–2759, 2009.

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

Luc Bauwens, Sébastien Laurent, and Jeroen VK Rombouts. Multivariate garch models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

Luca Benati and Paolo Surico. VAR analysis and the great moderation. *American Economic Review*, 99(4):1636–52, 2009.

Ahmed BenSaïda. Good and bad volatility spillovers: An asymmetric connectedness. *Journal of Financial Markets*, 43(3):78–95, 2019.

Stephen Beveridge and Charles R Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, 7(2):151–174, 1981.

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

Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics*, 72(3):498–505, 1990.

Trevor S Breusch. Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17(31):334–355, 1978.

Peter Bühlmann. Sieve bootstrap for smoothing in nonstationary time series. *Annals of Statistics*, 26(1):48–83, 1998.

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

Alessandro Casini. Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. *Journal of Econometrics*, forthcoming.

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

Likai Chen, Weining Wang, and Wei Biao Wu. Inference of breakpoints in high-dimensional time series. *Journal of the American Statistical Association*, forthcoming.

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

C-K Chu and James Stephen Marron. Comparison of two bandwidth selectors with dependent errors. *Annals of Statistics*, 19(4):1906–1918, 1991.

Timothy Cogley, Giorgio E Primiceri, and Thomas J Sargent. Inflation-gap persistence in the U.S. *American Economic Journal: Macroeconomics*, 2(1):43–69, 2010.

Rainer Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and Their Applications*, 62(1):139–168, 1996a.

Rainer Dahlhaus. On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and Their Applications*, 62(1):139–168, 1996b.

Rainer Dahlhaus. Locally stationary processes. In *Handbook of Statistics*, volume 30, pages 351–413. Elsevier, 2012.

Rainer Dahlhaus and Wolfgang Polonik. Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39, 2009.

Rainer Dahlhaus and Suhasini Subba Rao. Statistical inference for time-varying ARCH processes. *Annals of Statistics*, 34(3):1075–1114, 2006.

Rainer Dahlhaus, Stefan Richter, and Wei Biao Wu. Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2):1013–1044, 2019.

Holger Dette, Philip Preuß, and Mathias Vetter. A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association*, 106(495):1113–1124, 2011.

Francis X Diebold and Kamil Yilmaz. Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534): 158–171, 2009.

Paul Doukhan. *Mixing: properties and examples*, volume 85. Springer Science & Business Media, 2012.

Jianqing Fan and Tao Huang. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11(6):1031–1057, 2005.

Jianqing Fan, Mingjin Wang, and Qiwei Yao. Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society: series B*, 70 (4):679–702, 2008.

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

Marina Friedrich, Stephan Smeekes, and Jean-Pierre Urbain. Autoregressive wild bootstrap inference for nonparametric trends. *Journal of Econometrics*, 214(1):81–109, 2020.

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

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

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

Jiti Gao, Shin Kanaya, Degui Li, and Dag Tjøstheim. Uniform consistency for nonparametric estimation in null recurrent time series. *Econometric Theory*, 31(4):911–952, 2015.

Liudas Giraitis, George Kapetanios, and Tony Yates. Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, 179(1):46–65, 2014.

Liudas Giraitis, George Kapetanios, and Tony Yates. Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis*, 39(2):129–149, 2018.

Leslie G Godfrey. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, 46(6):1293–1301, 1978.

Peter Hall and Christopher C Heyde. *Martingale Limit Theory and Its Application.* Academic Press, 1980.

Peter G. Hall and Jeffrey S. Racine. Infinite order cross-validated local polynomial regression. *Journal of Econometrics*, 185(2):510–525, 2015.

James D Hamilton. *Time Series Analysis*, volume 2. Princeton University Press, New Jersey, 1994.

Bruce E Hansen. Strong laws for dependent heterogeneous processes. *Econometric theory*, 7(2):213–221, 1991.

Bruce E Hansen. Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica*, 60(4):967–972, 1992.

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

Bruce E Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748, 2008.

Atsushi Inoue and Lutz Kilian. Inference on impulse response functions in structural var models. *Journal of Econometrics*, 117(1):1–13, 2013.

Atsushi Inoue and Lutz Kilian. The uniform validity of impulse response inference in autoregressions. *Journal of Econometrics*, 215(3):450–472, 2020.

Thierry Jeantheau. Strong consistency of estimators for multivariate arch models. *Econometric theory*, 14(1):70–86, 1998.

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

Sayar Karmakar, Stefan Richter, and Wei Biao Wu. Simultaneous inference for time-varying models. *Journal of Econometrics*, forthcoming.

Lutz Kilian and Helmut Lütkepohl. *Structural Vector Autoregressive Analysis*. Cambridge University Press, 2017.

Degui Li, Peter C. B. Phillips, and Jiti Gao. Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression. *Econometric Theory*, 32(3): 655–685, 2016.

Degui Li, Peter C. B. Phillips, and Jiti Gao. Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics*, 215(2):607–632, 2020.

Qi Li and Jeffrey Racine. *Nonparametric Econometrics Theory and Practice*. Princeton University Press, New Jersey, 2007.

Shiqing Ling. Adaptive estimators and tests of stationary and nonstationary short-and long-memory arfima–garch models. *Journal of the American Statistical Association*, 98 (464):955–967, 2003.

Shiqing Ling and Michael McAleer. Asymptotic theory for a vector ARMA-GARCH model. *Econometric theory*, pages 280–310, 2003.

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

Tucker S McElroy and Anindya Roy. Model identification via total frobenius norm of multivariate spectra. *Journal of the Royal Statistical Society: Series B*, forthcoming, 2021.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

Franz C Palm, Stephan Smeekes, and Jean-Pierre Urbain. Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics*, 163(1):85–104, 2011.

Qunxing Pan, Xiaowen Mei, and Tianqing Gao. Modeling dynamic conditional correlations with leverage effects and volatility spillover effects: Evidence from the chinese and us stock markets affected by the recent trade friction. *The North American Journal of Economics and Finance*, 59:101591, 2022.

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

Youquan Pei, Tao Huang, and Jinhong You. Nonparametric fixed effects model for panel data with locally stationary regressors. *Journal of Econometrics*, 202(2):286–305, 2018.

Katerina Petrova. A quasi-Bayesian local likelihood approach to time varying parameter VAR models. *Journal of Econometrics*, 212(1):286–306, 2019.

Peter C. B. Phillips and Ji Hyung Lee. Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250–264, 2013.

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

Peter C. B. Phillips, Degui Li, and Jiti Gao. Estimating smooth structural change in cointergration models. *Journal of Econometrics*, 196(1):180–195, 2017.

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

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

Philip Preuss, Ruprecht Puchstein, and Holger Dette. Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association*, 110 (510):654–668, 2015.

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

Giorgio E Primiceri. Why inflation rose and fell: policy-makers' beliefs and us postwar stabilization policy. *Quarterly Journal of Economics*, 121(3):867–901, 2006.

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

Stefan Richter and Rainer Dahlhaus. Cross validation for locally stationary processes. *Annals of Statistics*, 47(4):2145–2173, 2019.

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

Xiaofeng Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235, 2010.

Christopher A Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

James H Stock and Mark W Watson. Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115, 2001.

James H Stock and Mark W Watson. Core inflation and trend inflation. *Review of Economics and Statistics*, 98(4):770–784, 2016a.

James H Stock and Mark W Watson. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier, 2016b.

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

Liangjun Su and Xia Wang. On time–varying factor models: estimation and testing. *Journal of Econometrics*, 198(1):84–101, 2017.

Y. Sun, Y. Hong, T. H. Lee, S. Wang, and X. Zhang. Time–varying model averaging. *Journal of Econometrics*, 222(2):974–992, 2021a.

Yuying Sun, Yongmiao Hong, Tae-Hwy Lee, Shouyang Wang, and Xinyu Zhang. Time-varying model averaging. *Journal of Econometrics*, 222(2):974–992, 2021b.

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

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

Michael Vogt. Nonparametric regression for locally stationary time series. *Annals of Statistics*, 40(5):2601–2633, 2012.

CS Withers. Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):477–480, 1981.

Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154, 2005.

Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436, 2004.

Wei Biao Wu and Zhou Zhou. Gaussian approximations for non-stationary multiple time series. *Statistica Sinica*, 21(3):1397–1413, 2011.

Yingcun Xia. Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society: Series B*, 60(4):797–811, 1998.

Zhuying Xu, Seonjin Kim, and Zhibiao Zhao. Locally stationary quantile regression for inflation and interest rates. *Journal of Business and Economic Statistics*, 39:Forthcoming, 2021.

K. Yousuf and S. Ng. Boosting high dimensional predictive regressions with time varying parameters. *Journal of Econometrics*, 224(1):60–87, 2020.

Bing Zhang and Xiao-Ming Li. Has there been any change in the comovement between the chinese and us stock markets? *International Review of Economics & Finance*, 29: 525–536, 2014.

Danna Zhang and Wei Biao Wu. Gaussian approximation for high dimensional time series. *Annals of Statistics*, 45(5):1895–1919, 2017.

Ting Zhang and Wei Biao Wu. Testing parametric assumptions of trends of a nonstationary time series. *Biometrika*, 98(3):599–614, 2011.

Ting Zhang and Wei Biao Wu. Inference of time-varying regression models. *Annals of Statistics*, 40(3):1376–1402, 2012.

Zhou Zhou and Wei Biao Wu. Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B*, 72(4):513–531, 2010.

Appendix A

Appendix for Chapter 2

In this appendix, I first present several preliminary lemmas in Appendix A.1, which are helpful to the development of the main results. I then prove the main results on the time–varying VMA(∞) process from Sections 2.2–2.3 in Appendix A.2. In what follows, M and $O(1)$ always stand for bounded constants, and may be different at each appearance.

A.1 Preliminary Lemmas

Lemma A.1.1. Suppose $\{Z_t, \mathscr{F}_t\}$ 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 A.1.1 is Corollary 3.1 of Hall and Heyde [1980].

Lemma A.1.2. Let $\{Z_t, \mathscr{F}_t\}$ be a martingale difference sequence. Suppose that $|Z_t| \leq M$ for a constant M, $t = 1, \ldots, T$. Let $V_T = \sum_{t=1}^T \mathrm{Var}\,(Z_t|\mathscr{F}_{t-1}) \leq V$ 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 A.1.2 is Proposition 2.1 of Freedman [1975].

Lemma A.1.3. The following algebraic decompositions hold true.

1. $\mathbb{B}_t(L) = \sum_{j=0}^\infty \mathbf{B}_{j,t} L^j$ can be decomposed as $\mathbb{B}_t(L) = \mathbb{B}_t(1) - (1-L)\widetilde{\mathbb{B}}_t(L)$, where $\widetilde{\mathbb{B}}_t(L) = \sum_{j=0}^\infty \widetilde{\mathbf{B}}_{j,t} L^j$ and $\widetilde{\mathbf{B}}_{j,t} = \sum_{k=j+1}^\infty \mathbf{B}_{k,t}$.

2. $\mathbb{B}_t^r(L) = \sum_{j=0}^\infty (\mathbf{B}_{j+r,t} \otimes \mathbf{B}_{j,t}) L^j$ can be decomposed as $\mathbb{B}_t^r(L) = \mathbb{B}_t^r(1) - (1-L)\widetilde{\mathbb{B}}_t^r(L)$, where $\widetilde{\mathbb{B}}_t^r(L) = \sum_{j=0}^\infty \widetilde{\mathbf{B}}_{j,t}^r L^j$ and $\widetilde{\mathbf{B}}_{j,t}^r = \sum_{k=j+1}^\infty (\mathbf{B}_{k+r,t} \otimes \mathbf{B}_{k,t})$.

In addition, let Assumption 2.2.1 hold, then

3. $\max_{t\geq 1} \sum_{j=0}^\infty |\widetilde{\mathbf{B}}_{j,t}| < \infty$;

4. $\limsup_{T\to\infty} \sum_{t=1}^{T-1} |\widetilde{\mathbb{B}}_{t+1}(1) - \widetilde{\mathbb{B}}_t(1)| < \infty$;

5. $\max_{t\geq 1} \sum_{j=0}^\infty |\widetilde{\mathbf{B}}_{j,t}^r| < \infty$;

6. $\max_{t\geq 1} \sum_{r=1}^\infty |\widetilde{\mathbb{B}}_t^r(1)| < \infty$;

7. $\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{r=0}^{\infty}|\widetilde{\mathbb{B}}_{t+1}^{r}(1)-\widetilde{\mathbb{B}}_{t}^{r}(1)|<\infty$.

Lemma A.1.4. Let Assumptions 2.2.1 and 2.2.2 hold, and let $\{\mathbf{W}_{T,t}(\cdot)\}_{t=1}^{T}$ be a sequence of $m\times d$ matrices of functions, where $m\geq 1$ is fixed, and each functional component is Lipschitz continuous and defined on a compact set $[a,b]$. Moreover, suppose that (1) $\sup_{\tau\in[a,b]}\sum_{t=1}^{T}|\mathbf{W}_{T,t}(\tau)|=O(1)$, and (2) $T^{\frac{2}{\delta}}d_T\log T\to 0$, where $d_T=\sup_{\tau\in[a,b],t\geq 1}|\mathbf{W}_{T,t}(\tau)|$. As $T\to\infty$,

$$\sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}(\tau)\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t\right|=O_P\left(\sqrt{d_T\log T}\right).$$

Lemma A.1.5. Let the conditions of Lemma A.1.4 hold. Suppose further $T^{\frac{4}{\delta}}d_T\log T\to 0$, $\max_{t\geq 1}E[\|\boldsymbol{\varepsilon}_t\|^4|\mathscr{F}_{t-1}]<\infty$ a.s. and $\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}|\mathbf{W}_{T,t+1}(\tau)-\mathbf{W}_{T,t}(\tau)|=O(d_T)$. As $T\to\infty$

1. $\sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T}\left(\mathbf{I}_d\otimes\mathbf{W}_{T,t}(\tau)\right)\mathbb{B}_t^{0}(1)\left(\mathrm{vec}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top]-\mathrm{vec}[\mathbf{I}_d]\right)\right|=O_P\left(\sqrt{d_T\log T}\right)$;

2. $\sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T}\left(\mathbf{I}_d\otimes\mathbf{W}_{T,t}(\tau)\right)\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|=O_P\left(\sqrt{d_T\log T}\right)$;

where $\boldsymbol{\zeta}_t=\sum_{r=1}^{\infty}\sum_{s=0}^{\infty}\{\mathbf{B}_{s+r,t}\boldsymbol{\varepsilon}_{t-r}\}\otimes\mathbf{B}_{s,t}$.

Lemma A.1.6. Under Assumption 2.4.1, there exists a time–varying VMA(∞) process

$$\widetilde{\mathbf{y}}_t=\boldsymbol{\mu}(\tau_t)+\sum_{j=0}^{\infty}\mathbf{D}_j^{\boldsymbol{\varepsilon}}(\tau_t)\boldsymbol{\varepsilon}_{t-j}+\sum_{j=0}^{\infty}\mathbf{D}_j^{\mathbf{v}}(\tau_t)\mathbf{v}_{t-j}$$

such that $\max_{t\geq 1}\|\mathbf{y}_t-\widetilde{\mathbf{y}}_t\|_\delta=O(T^{-1})$, where

$$\begin{aligned}
\boldsymbol{\mu}(\tau) &= \sum_{j=0}^{\infty}\sum_{l=0}^{q}\boldsymbol{\Psi}_j(\tau)\mathbf{B}_l(\tau)\mathbf{g}(\tau),\quad \boldsymbol{\Psi}_j(\tau)=\mathbf{J}\boldsymbol{\Phi}^j(\tau)\mathbf{J}^\top,\\
\boldsymbol{\Phi}(\tau) &= \begin{pmatrix}\mathbf{A}_1(\tau) & \cdots & \mathbf{A}_{p-1}(\tau) & \mathbf{A}_p(\tau)\\ \mathbf{I}_d & \cdots & \mathbf{0}_d & \mathbf{0}_d\\ \vdots & \ddots & \vdots & \vdots\\ \mathbf{0}_d & \cdots & \mathbf{I}_d & \mathbf{0}_d\end{pmatrix},\quad \mathbf{J}=\left[\mathbf{I}_d,\mathbf{0}_{d\times d(p-1)}\right],\\
\mathbf{D}_j^{\boldsymbol{\varepsilon}}(\tau) &= \boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}(\tau),\quad \mathbf{D}_j^{\mathbf{v}}(\tau)=\sum_{b=\max(0,j-q)}^{j}\mathbf{D}_{b,j-b}^{\mathbf{v}}(\tau),\\
\mathbf{D}_{j,l}^{\mathbf{v}}(\tau) &= \sum_{k=0}^{j}\boldsymbol{\Psi}_k(\tau)\mathbf{B}_l(\tau)\mathbf{C}_{j-k}(\tau).
\end{aligned}$$

Moreover, $\widetilde{\mathbf{y}}_t$ and $\mathbf{x}_t$ admit the following expression

$$\begin{pmatrix}\widetilde{\mathbf{y}}_t\\ \mathbf{x}_t\end{pmatrix}=\begin{pmatrix}\boldsymbol{\mu}(\tau_t)\\ \mathbf{g}(\tau_t)\end{pmatrix}+\sum_{j=0}^{\infty}\mathbf{D}_j(\tau_t)\begin{pmatrix}\boldsymbol{\varepsilon}_{t-j}\\ \mathbf{v}_{t+1-j}\end{pmatrix},$$

where $\mathbf{D}_j(\tau)=\begin{pmatrix}\mathbf{D}_j^{\boldsymbol{\varepsilon}}(\tau) & \mathbf{D}_{j-1}^{\mathbf{v}}(\tau)\\ \mathbf{0} & \mathbf{C}_{j-1}(\tau)\end{pmatrix}$, and $\mathbf{D}_j^{\mathbf{v}}(\tau)=0$ and $\mathbf{C}_j(\tau)=0$ for $j<0$. Here, $\mathbf{D}_j(\cdot)$'s satisfy the same conditions as those in Assumption 2.3.1.

Lemma A.1.7. Suppose Assumptions 2.3.2–2.4.2 hold. As $T\to\infty$,

1. $\sup_{\tau\in[h,1-h]}\left|\frac{1}{T}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^\top\mathbf{K}_\tau\mathbf{X}_{\widetilde{\mathbf{C}},\tau}-\mathbf{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}(\tau)\otimes\mathbf{\Lambda}_1\right|=O_P\left(h^2+\sqrt{\frac{\log T}{Th}}\right)$ with $\mathbf{\Lambda}_1=\mathrm{diag}(\tilde{c}_0,\tilde{c}_2)$;
2. $\sup_{\tau\in[h,1-h]}\left|\frac{1}{T}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^\top\mathbf{K}_\tau\mathbf{X}_{\mathbf{C},\tau}-\mathbf{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}^\top(\tau)\otimes\mathbf{\Lambda}_2\right|=O_P\left(h^2+\sqrt{\frac{\log T}{Th}}\right)$ with $\mathbf{\Lambda}_2=[\tilde{c}_0,0]^\top$;
3. $\sup_{\tau\in[0,1]}\left|\frac{1}{T}\mathbf{Z}_\tau\mathbf{K}_\tau\boldsymbol{\eta}\right|=O_P\left(\sqrt{\frac{\log T}{Th}}\right)$, where $\boldsymbol{\eta}=(\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_T)^\top$.

Lemma A.1.8. Suppose Assumptions 2.3.2–2.4.2 hold. As $T\to\infty$,

1. $\frac{1}{T}\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT}-\mathbf{S})^\top(\mathbf{I}_{dT}-\mathbf{S})\mathbf{X}_{\mathbf{C}}=\mathbf{\Sigma}+o_P(1)$,
   where $\mathbf{\Sigma}=\int_0^1\mathbf{\Sigma}_{\mathbf{X}_{\mathbf{C}}}(\tau)\mathrm{d}\tau-\int_0^1\mathbf{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau)\mathbf{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}^{-1}(\tau)\mathbf{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}^\top(\tau)\mathrm{d}\tau$;
2. $\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT}-\mathbf{S})^\top(\mathbf{I}_{dT}-\mathbf{S})\widetilde{\mathbf{X}}=o_P(\sqrt{T})$, where $\widetilde{\mathbf{X}}=([\mathbf{X}_{\widetilde{\mathbf{C}},1}^\top\boldsymbol{\theta}(\tau_1)]^\top,\ldots,[\mathbf{X}_{\widetilde{\mathbf{C}},T}^\top\boldsymbol{\theta}(\tau_T)]^\top)^\top$.

A.2 Proofs of the Main Results

Proof of Lemma 2.2.1.

By the BN decomposition in Lemma A.1.3, I have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}(\mathbf{x}_t-E(\mathbf{x}_t))=\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t+\frac{1}{\sqrt{T}}\widetilde{\mathbb{B}}_1(L)\boldsymbol{\varepsilon}_0-\frac{1}{\sqrt{T}}\widetilde{\mathbb{B}}_{\lfloor Tr\rfloor}(L)\boldsymbol{\varepsilon}_{\lfloor Tr\rfloor}$$
$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor-1}\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t:=\mathbf{I}_{T,1}+\mathbf{I}_{T,2}+\mathbf{I}_{T,3}+\mathbf{I}_{T,4}.$$

By the usual functional central limit theory for martingale difference sequences, I have

$$\mathbf{\Sigma}^{-1/2}(r)\mathbf{I}_{T,1}\to_D\mathbf{W}(r).$$

In addition, I have $\mathbf{I}_{T,2}\to_P 0$ uniformly over $r\in[0,1]$ as $\left\|\widetilde{\mathbb{B}}_1(L)\boldsymbol{\varepsilon}_0\right\|_1<\infty$ by Lemma A.1.3.

For $\mathbf{I}_{T,3}$, I need to show that

$$\sup_{r\in[0,1]}\left|\frac{1}{\sqrt{T}}\widetilde{\mathbb{B}}_{\lfloor Tr\rfloor}(L)\boldsymbol{\varepsilon}_{\lfloor Tr\rfloor}\right|\to_P 0$$

which holds if $\max_{1\le t\le T}T^{-1}|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_t|^2\to_P 0$. This is equivalent to show for any $\nu>0$

$$\frac{1}{T}\sum_{t=1}^{T}E\left[|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_t|^2I(|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_t|^2>T\nu)\right]\to 0.$$

Similar to the proofs of Lemma 2.2.4, this is satisfied due to

$$\|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_t\|_\delta\le M\sum_{j=1}^{\infty}|\widetilde{\mathbf{B}}_{j,t}|<\infty.$$

Finally, for $I_{T,4}$, as $\sum_{t=1}^{T-1}\left\|\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t\right\|_1<\infty$ by Lemma A.1.3, I have

$$\begin{aligned}\sup_{r\in[0,1]}|I_{T,4}| &\leq \sup_{r\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor-1}\left|\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t\right| \\ &\leq \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left|\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t\right| = O_P(1/\sqrt{T}).\end{aligned}$$

The proof is now completed.

□

Proof of Proposition 2.2.1.

(1). Start from Example 1. Let ρ denote the largest eigenvalue of $\mathbf{\Phi}_t$ uniformly over t. Then, $\rho<1$ by the condition in Proposition 2.2.1. Similar to the proof of Proposition 2.4 in Dahlhaus and Polonik [2009], I have $\max_{t\geq 1}|\prod_{i=0}^{j-1}\mathbf{\Phi}_{t-i}|\leq M\rho^j$, which yields that

$$\max_{t\geq 1}\sum_{j=1}^{\infty}j\,|\mathbf{B}_{j,t}| = \max_{t\geq 1}\sum_{j=1}^{\infty}j\left|\mathbf{J}\prod_{i=0}^{j-1}\mathbf{\Phi}_{t-i}\mathbf{J}^\top\right| \leq M\sum_{j=1}^{\infty}j\rho^j = O(1).$$

In addition, for any conformable matrices $\{\mathbf{A}_i\}$ and $\{\mathbf{B}_i\}$, since

$$\prod_{i=1}^{r}\mathbf{A}_i-\prod_{i=1}^{r}\mathbf{B}_i=\sum_{j=1}^{r}\left(\prod_{k=1}^{j-1}\mathbf{A}_k\right)(\mathbf{A}_j-\mathbf{B}_j)\left(\prod_{k=j+1}^{r}\mathbf{B}_k\right),$$

I then obtain that

$$\begin{aligned}&\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\left|\mathbf{J}\left(\prod_{i=0}^{j-1}\mathbf{\Phi}_{t+1-i}-\prod_{i=0}^{j-1}\mathbf{\Phi}_{t-i}\right)\mathbf{J}^\top\right| \\ =\;&\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\left|\mathbf{J}\sum_{m=1}^{j}\left(\prod_{k=1}^{m-1}\mathbf{\Phi}_{t+2-k}\right)(\mathbf{\Phi}_{t+2-m}-\mathbf{\Phi}_{t+1-m})\left(\prod_{k=m}^{j}\mathbf{\Phi}_{t+1-k}\right)\mathbf{J}^\top\right| \\ \leq\;& M\sum_{j=1}^{\infty}j^2\rho^{j-1}\limsup_{T\to\infty}\sum_{t=1}^{T-1}|\mathbf{\Phi}_{t+1}-\mathbf{\Phi}_t| = O(1)\end{aligned}$$

given the condition in Proposition 2.2.1.

Consider Example 2. Similar to Example 1,

$$\max_{t\geq 1}\sum_{b=1}^{\infty}b\,|\mathbf{D}_{b,t}|\leq M\max_{t\geq 1}\sum_{b=1}^{\infty}b\sum_{j=b-q}^{b}|\mathbf{B}_{j,t}|\leq M\sum_{b=1}^{\infty}b\rho^b=O(1).$$

In addition,

$$\begin{aligned}
&\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty} b\,|\mathbf{D}_{b,t+1}-\mathbf{D}_{b,t}| \\
\leq\ & \limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty} b \sum_{j=\max(0,b-q)}^{b} |\mathbf{B}_{j,t+1}-\mathbf{B}_{j,t}|\cdot|\mathbf{\Theta}_{b-j,t+1-j}| \\
&+\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty} b \sum_{j=\max(0,b-q)}^{b} |\mathbf{B}_{j,t}|\cdot|\mathbf{\Theta}_{b-j,t+1-j}-\mathbf{\Theta}_{b-j,t-j}| \\
\leq\ & \max_{m,t}|\mathbf{\Theta}_{m,t}|\cdot q\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty} b\,|\mathbf{B}_{b,t+1}-\mathbf{B}_{b,t}| \\
&+\left(\max_{t\geq 1}\sum_{b=1}^{\infty} b\sum_{j=\max(0,b-q)}^{b}|\mathbf{B}_{j,t}|\right)\cdot\left(\max_{m}\limsup_{T\to\infty}\sum_{t=1}^{T-1}|\mathbf{\Theta}_{m,t+1}-\mathbf{\Theta}_{m,t}|\right)=O(1).
\end{aligned}$$

(2). By the proof of part (1) and the condition of Proposition 2.2.1, it suffices to show that $\max_{t\geq 1}\sum_{j=1}^{\infty} j\,|\mathbf{D}_{j,t}|<\infty$ and $\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\,|\mathbf{D}_{j,t+1}-\mathbf{D}_{j,t}|<\infty$. Write

$$\begin{aligned}
&\max_{t\geq 1}\sum_{j=1}^{\infty} j\,|\mathbf{D}_{j,t}| \leq M\max_{t\geq 1}\sum_{j=1}^{\infty} j\sum_{k=0}^{j}|\mathbf{B}_{k,t}|\cdot|\mathbf{C}_{j-k,t-k}| \\
=\ & M\max_{t\geq 1}\sum_{k=0}^{\infty}\sum_{j=k}^{\infty} j\,|\mathbf{B}_{k,t}|\cdot|\mathbf{C}_{j-k,t-k}| = M\max_{t\geq 1}\sum_{k=0}^{\infty}|\mathbf{B}_{k,t}|\sum_{j=0}^{\infty}(k+j)\,|\mathbf{C}_{j,t-k}| \\
=\ & M\max_{t\geq 1}\sum_{j=0}^{\infty} j\,|\mathbf{B}_{j,t}|\sum_{k=1}^{\infty}|\mathbf{C}_{k,t-j}| + M\max_{t\geq 1}\sum_{k=0}^{\infty}|\mathbf{B}_{k,t}|\sum_{j=0}^{\infty} j\,|\mathbf{C}_{j,t-k}| = O(1).
\end{aligned}$$

In addition,

$$\begin{aligned}
&\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\,|\mathbf{D}_{j,t+1}-\mathbf{D}_{j,t}| \leq \limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\sum_{k=0}^{j}|\mathbf{B}_{k,t}|\cdot|\mathbf{\Theta}_{t-k}|\cdot|\mathbf{C}_{j-k,t+1-k}-\mathbf{C}_{j-k,t-k}| \\
&+\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\sum_{k=0}^{j}|\mathbf{B}_{k,t}|\cdot|\mathbf{\Theta}_{t+1-k}-\mathbf{\Theta}_{t-k}|\cdot|\mathbf{C}_{j-k,t+1-k}| \\
&+\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\sum_{k=0}^{j}|\mathbf{B}_{k,t+1}-\mathbf{B}_{k,t}|\cdot|\mathbf{\Theta}_{t+1-k}|\cdot|\mathbf{C}_{j-k,t+1-k}| := I_{T,1}+I_{T,2}+I_{T,3}.
\end{aligned}$$

I only show that $I_{T,1}$ is bounded below, as the proofs of $I_{T,2}$ and $I_{T,3}$ can be established similarly.

$$\begin{aligned}
I_{T,1} &\leq \max_{t\geq 1}|\mathbf{\Theta}_t|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\sum_{k=0}^{j}|\mathbf{B}_{k,t}|\cdot|\mathbf{C}_{j-k,t+1-k}-\mathbf{C}_{j-k,t-k}| \\
&= \max_{t\geq 1}|\mathbf{\Theta}_t|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{k=0}^{\infty}\sum_{j=k}^{\infty}j\,|\mathbf{B}_{k,t}|\cdot|\mathbf{C}_{j-k,t+1-k}-\mathbf{C}_{j-k,t-k}| \\
&= \max_{t\geq 1}|\mathbf{\Theta}_t|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{k=0}^{\infty}|\mathbf{B}_{k,t}|\sum_{j=0}^{\infty}(j+k)\,|\mathbf{C}_{j,t+1-k}-\mathbf{C}_{j,t-k}| \\
&\leq \max_{t\geq 1}|\mathbf{\Theta}_t|\cdot\max_{t\geq 1}\sum_{k=0}^{\infty}|\mathbf{B}_{k,t}|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=0}^{\infty}j\,|\mathbf{C}_{j,t+1-k}-\mathbf{C}_{j,t-k}| \\
&\quad+\max_{t\geq 1}|\mathbf{\Theta}_t|\cdot\max_{t\geq 1}\sum_{k=0}^{\infty}k\,|\mathbf{B}_{k,t}|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=0}^{\infty}|\mathbf{C}_{j,t+1-k}-\mathbf{C}_{j,t-k}| = O(1).
\end{aligned}$$

The proof is now completed. □

Proof of Lemma 2.2.2.

By Lemma A.1.3, I have

$$\mathbf{x}_t = \boldsymbol{\mu}_t + \mathbb{B}_t(1)\boldsymbol{\varepsilon}_t + \widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_{t-1} - \widetilde{\mathbb{B}}_t(L)\boldsymbol{\varepsilon}_t,$$

which yields

$$\begin{aligned}
&\sum_{t=1}^{T}\mathbf{W}_{T,t}\left(\mathbf{x}_t - E(\mathbf{x}_t)\right) = \sum_{t=1}^{T}\mathbf{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t + \mathbf{W}_{T,1}\widetilde{\mathbb{B}}_1(L)\boldsymbol{\varepsilon}_0 - \mathbf{W}_{T,T}\widetilde{\mathbb{B}}_T(L)\boldsymbol{\varepsilon}_T \\
&+\sum_{t=1}^{T-1}\left(\mathbf{W}_{T,t+1}\widetilde{\mathbb{B}}_{t+1}(L) - \mathbf{W}_{T,t}\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t := \mathbf{I}_{T,1}+\mathbf{I}_{T,2}+\mathbf{I}_{T,3}+\mathbf{I}_{T,4}.
\end{aligned}$$

For $\mathbf{I}_{T,1}$, by Assumption 2.2.2, I have

$$\begin{aligned}
&E\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t\right|^2 = \operatorname{tr}\left(\sum_{t=1}^{T}\mathbf{W}_{T,t}\mathbb{B}_t(1)E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top)\mathbb{B}_t^\top(1)\mathbf{W}_{T,t}^\top\right) \\
&\leq M\sum_{t=1}^{T}|\mathbf{W}_{T,t}|^2 \leq M\max_{t\geq 1}|\mathbf{W}_{T,t}|\sum_{t=1}^{T}|\mathbf{W}_{T,t}| = O(d_T).
\end{aligned}$$

Hence, $|\mathbf{I}_{T,1}| = O_P(\sqrt{d_T})$.

Also, $|\mathbf{I}_{T,2}| = O_P(d_T)$ and $|\mathbf{I}_{T,3}| = O_P(d_T)$, since $\max_{t\geq 1}|\mathbf{W}_{T,t}| = O(d_T)$, $\|\widetilde{\mathbb{B}}_1(L)\boldsymbol{\varepsilon}_0\|_1 < \infty$ and $\|\widetilde{\mathbb{B}}_T(L)\boldsymbol{\varepsilon}_t\|_1 < \infty$ by Lemma A.1.3.

For $\mathbf{I}_{T,4}$,

$$\begin{aligned}
&\sum_{t=1}^{T-1}\left(\mathbf{W}_{T,t+1}\widetilde{\mathbb{B}}_{t+1}(L) - \mathbf{W}_{T,t}\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t \\
&= \sum_{t=1}^{T-1}\left(\mathbf{W}_{T,t+1}-\mathbf{W}_{T,t}\right)\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t + \sum_{t=1}^{T-1}\mathbf{W}_{T,t}\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t. \qquad \text{(A.2.1)}
\end{aligned}$$

Note that for the first term on the right hand side of (A.2.1),

$$\left\|\sum_{t=1}^{T-1} (\mathbf{W}_{T,t+1} - \mathbf{W}_{T,t})\, \widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t\right\|_1 \le \max_{t\ge 1}\left\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t\right\|_1 \cdot \sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1} - \mathbf{W}_{T,t}| = O(d_T)$$

by Lemma A.1.3 and the conditions on $\mathbf{W}_{T,t}$. For the second term on the right hand side of (A.2.1), write

$$\begin{aligned}
&\left\|\sum_{t=1}^{T-1} \mathbf{W}_{T,t}\left(\widetilde{\mathbb{B}}_{t+1}(L) - \widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t\right\|_1 \\
\le\ & \max_{t\ge 1}\|\boldsymbol{\varepsilon}_t\|_1 \cdot \max_{t\ge 1}|\mathbf{W}_{T,t}| \sum_{t=1}^{T-1} |\widetilde{\mathbb{B}}_{t+1}(1) - \widetilde{\mathbb{B}}_t(1)| \\
\le\ & M \max_{t\ge 1}|\mathbf{W}_{T,t}| \sum_{t=1}^{T-1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty} |\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}| \\
=\ & M \max_{t\ge 1}|\mathbf{W}_{T,t}| \sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\,|\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}| = O(d_T).
\end{aligned}$$

Thus, I have proved that $|\sum_{t=1}^T \mathbf{W}_{T,t}(\mathbf{x}_t - E(\mathbf{x}_t))| = O_P(\sqrt{d_T})$.

I now prove $|\sum_{t=1}^T \mathbf{W}_{T,t}\left(\mathbf{x}_t\mathbf{x}_{t+p}^\top - E\left(\mathbf{x}_t\mathbf{x}_{t+p}^\top\right)\right)| = O_P(\sqrt{d_T})$. Start from $p = 0$ and write

$$\begin{aligned}
\mathbf{x}_t\mathbf{x}_t^\top =\ & \boldsymbol{\mu}_t\boldsymbol{\mu}_t^\top + \boldsymbol{\mu}_t\sum_{j=0}^{\infty}\boldsymbol{\varepsilon}_{t-j}^\top\mathbf{B}_{j,t}^\top + \sum_{j=0}^{\infty}\mathbf{B}_{j,t}\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\mu}_t^\top + \sum_{j=0}^{\infty}\mathbf{B}_{j,t}\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j}^\top\mathbf{B}_{j,t}^\top \\
&+ \sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\mathbf{B}_{j,t}\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j-r}^\top\mathbf{B}_{j+r,t}^\top + \sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\mathbf{B}_{j+r,t}\boldsymbol{\varepsilon}_{t-j-r}\boldsymbol{\varepsilon}_{t-j}^\top\mathbf{B}_{j,t}^\top,
\end{aligned}$$

which yields

$$\begin{aligned}
&\text{vec}\left[\mathbf{W}_{T,t}\left(\mathbf{x}_t\mathbf{x}_t^\top - E\left(\mathbf{x}_t\mathbf{x}_t^\top\right)\right)\right] \\
=\ & (\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{j=0}^{\infty}(\mathbf{B}_{j,t}\otimes\boldsymbol{\mu}_t)\,\boldsymbol{\varepsilon}_{t-j} + (\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{j=0}^{\infty}(\boldsymbol{\mu}_t\otimes\mathbf{B}_{j,t})\,\boldsymbol{\varepsilon}_{t-j} \\
&+ (\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{j=0}^{\infty}(\mathbf{B}_{j,t}\otimes\mathbf{B}_{j,t})\,\text{vec}[\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j}^\top - \mathbf{I}_d] \\
&+ (\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\mathbf{B}_{j+r,t}\otimes\mathbf{B}_{j,t})\text{vec}[\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j-r}^\top] \\
&+ (\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\mathbf{B}_{j,t}\otimes\mathbf{B}_{j+r,t})\text{vec}[\boldsymbol{\varepsilon}_{t-j-r}\boldsymbol{\varepsilon}_{t-j}^\top].
\end{aligned}$$

Consequently, I obtain

$$
\begin{aligned}
&\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}\left(\mathbf{x}_t\mathbf{x}_t^\top - E\left(\mathbf{x}_t\mathbf{x}_t^\top\right)\right)\right| \leq 2\left|\sum_{t=1}^{T}(\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{j=0}^{\infty}(\boldsymbol{\mu}_t\otimes\mathbf{B}_{j,t})\,\boldsymbol{\varepsilon}_{t-j}\right| \\
&+\left|\sum_{t=1}^{T}(\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{j=0}^{\infty}(\mathbf{B}_{j,t}\otimes\mathbf{B}_{j,t})\,\mathrm{vec}[\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j}^\top - \mathbf{I}_d]\right| \\
&+2\left|\sum_{t=1}^{T}(\mathbf{I}_d\otimes\mathbf{W}_{T,t})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\mathbf{B}_{j+r,t}\otimes\mathbf{B}_{j,t})\mathrm{vec}[\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j-r}^\top]\right| := I_{T,5}+I_{T,6}+I_{T,7}.
\end{aligned}
$$

By the development of $\sum_{t=1}^{T}\mathbf{W}_{T,t}\,(\mathbf{x}_t - E(\mathbf{x}_t))$, it is easy to know that $I_{T,5}$ is $O_P(\sqrt{d_T})$. For $I_{T,6}$, by Lemma A.1.3, write

$$
\begin{aligned}
I_{T,6} \leq &\left|\sum_{t=1}^{T}(\mathbf{I}_d\otimes\mathbf{W}_{T,t})\,\mathbb{B}_t^0(1)\left(\mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top)-\mathrm{vec}(\mathbf{I}_d)\right)\right| \\
&+\left|(\mathbf{I}_d\otimes\mathbf{W}_{T,1})\,\widetilde{\mathbb{B}}_1^0(L)\mathrm{vec}(\boldsymbol{\varepsilon}_0\boldsymbol{\varepsilon}_0^\top)\right| + \left|(\mathbf{I}_d\otimes\mathbf{W}_{T,T})\,\widetilde{\mathbb{B}}_T^0(L)\mathrm{vec}(\boldsymbol{\varepsilon}_T\boldsymbol{\varepsilon}_T^\top)\right| \\
&+\left|\sum_{t=1}^{T-1}\left((\mathbf{I}_d\otimes\mathbf{W}_{T,t+1})\,\widetilde{\mathbb{B}}_{t+1}^0(L)-(\mathbf{I}_d\otimes\mathbf{W}_{T,t})\,\widetilde{\mathbb{B}}_t^0(L)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right| \\
:= &\; I_{T,61}+I_{T,62}+I_{T,63}+I_{T,64}.
\end{aligned}
$$

Let $\mathbf{Z}_t = \mathrm{vec}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top - \mathbf{I}_d)$ for notational simplicity. For $I_{T,61}$, write

$$
\begin{aligned}
&\left\|\sum_{t=1}^{T}(\mathbf{I}_d\otimes\mathbf{W}_{T,t})\,\mathbb{B}_t^0(1)\mathbf{Z}_t\right\|^2 \\
\leq\;& M\left(\max_{t\geq 1}\sum_{j=0}^{\infty}|\mathbf{B}_{j,t}|^2\right)^2\sum_{t=1}^{T}|\mathbf{W}_{T,t}|^2\,\|\mathbf{Z}_t\|^2 \\
\leq\;& M\max_{t\geq 1}|\mathbf{W}_{T,t}|\sum_{t=1}^{T}|\mathbf{W}_{T,t}| = O(d_T),
\end{aligned}
$$

which implies that $I_{T,61} = O_P(\sqrt{d_T})$. Similar to the proof of $\mathbf{I}_{T,2}$ and $\mathbf{I}_{T,3}$, I can prove that $I_{T,62}$ and $I_{T,63}$ are $O_P(d_T)$. For $I_{T,64}$, I have

$$
\begin{aligned}
I_{T,64} \leq &\left|\sum_{t=1}^{T-1}(\mathbf{I}_d\otimes(\mathbf{W}_{T,t+1}-\mathbf{W}_{T,t}))\,\widetilde{\mathbb{B}}_{t+1}^0(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right| \\
&+\left|\sum_{t=1}^{T-1}(\mathbf{I}_d\otimes\mathbf{W}_{T,t})\left(\widetilde{\mathbb{B}}_{t+1}^0(L)-\widetilde{\mathbb{B}}_t^0(L)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right|.
\end{aligned}
$$

Similar to the proof of $\mathbf{I}_{T,4}$, by Lemma A.1.3, I can prove that $I_{T,64}$ is $O_P(d_T)$. Then I can conclude that $I_{T,6} = O_P(\sqrt{d_T})$.

For $I_{T,7}$, using Lemma A.1.3, I have

$$\begin{aligned} I_{T,7} \leq & \left| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}) \sum_{r=1}^{\infty} \mathbb{B}_t^r(1) \mathrm{vec}\left(\varepsilon_t \varepsilon_{t-r}^\top\right)\right| + \left|(\mathbf{I}_d \otimes \mathbf{W}_{T,1}) \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_1^r(L) \mathrm{vec}\left(\varepsilon_0 \varepsilon_{-r}^\top\right)\right| \\ & + \left|(\mathbf{I}_d \otimes \mathbf{W}_{T,T}) \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_T^r(L) \mathrm{vec}\left(\varepsilon_T \varepsilon_{T-r}^\top\right)\right| \\ & + \left| \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \left((\mathbf{I}_d \otimes \mathbf{W}_{T,t+1}) \widetilde{\mathbb{B}}_{t+1}^r(L) - (\mathbf{I}_d \otimes \mathbf{W}_{T,t}) \widetilde{\mathbb{B}}_t^r(L)\right) \mathrm{vec}\left(\varepsilon_t \varepsilon_{t-r}^\top\right)\right| \\ := & \ I_{T,71} + I_{T,72} + I_{T,73} + I_{T,74}. \end{aligned}$$

For $I_{T,71}$, by Lemma A.1.3, I further write

$$\begin{aligned} & \left\| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}) \sum_{r=1}^{\infty} \mathbb{B}_t^r(1) \mathrm{vec}\left(\varepsilon_t \varepsilon_{t-r}^\top\right)\right\|^2 \\ = & \ E \,\mathrm{tr} \left\{ \sum_{t=1}^{T} \sum_{s=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}) \sum_{r,k=1}^{\infty} \mathbb{B}_t^r(1) \mathrm{vec}\left(\varepsilon_t \varepsilon_{t-r}^\top\right) \mathrm{vec}^\top\left(\varepsilon_s \varepsilon_{s-k}^\top\right) \mathbb{B}_s^{k,\top}(1)(\mathbf{I}_d \otimes \mathbf{W}_{T,s}^\top)\right\} \\ \leq & \ M \sum_{t=1}^{T} |\mathbf{W}_{T,t}|^2 \sum_{r=1}^{\infty} |\mathbb{B}_t^r(1)|^2 \leq M \left(\max_{t\geq 1} \sum_{r=1}^{\infty} |\mathbb{B}_t^r(1)|\right)^2 \max_{t\geq 1} |\mathbf{W}_{T,t}| \sum_{t=1}^{T} |\mathbf{W}_{T,t}| = O(d_T). \end{aligned}$$

In addition, similar to the proof of $\mathbf{I}_{T,2}$ to $\mathbf{I}_{T,4}$, I can show that $I_{T,72}$ to $I_{T,74}$ are $O_P(d_T)$.

Combining the above results, I have proved the case of $p = 0$.

Similar to the development of $p = 0$, I can consider the case with $p \geq 1$ given p is a fixed number. The details are omitted due to similarity. The proof is now completed. □

Proof of Lemma 2.2.3.

(1). By Lemma A.1.3, I have

$$\mathbf{x}_t = \boldsymbol{\mu}_t + \mathbb{B}_t(1)\varepsilon_t + \widetilde{\mathbb{B}}_t(L)\varepsilon_{t-1} - \widetilde{\mathbb{B}}_t(L)\varepsilon_t.$$

I are then able to write

$$\begin{aligned} & \sup_{\tau\in[a,b]} \left| \sum_{t=1}^{T} \mathbf{W}_{T,t}(\tau)(\mathbf{x}_t - E(\mathbf{x}_t))\right| \\ \leq & \sup_{\tau\in[a,b]} \left| \sum_{t=1}^{T} \mathbf{W}_{T,t}(\tau)\mathbb{B}_t(1)\varepsilon_t\right| + \sup_{\tau\in[a,b]} \left|\mathbf{W}_{T,1}(\tau)\widetilde{\mathbb{B}}_1(L)\varepsilon_0\right| + \sup_{\tau\in[a,b]} \left|\mathbf{W}_{T,T}(\tau)\widetilde{\mathbb{B}}_T(L)\varepsilon_T\right| \\ & + \sup_{\tau\in[a,b]} \left| \sum_{t=1}^{T-1} \left(\mathbf{W}_{T,t+1}(\tau)\widetilde{\mathbb{B}}_{t+1}(L) - \mathbf{W}_{T,t}(\tau)\widetilde{\mathbb{B}}_t(L)\right)\varepsilon_t\right| := I_{T,1} + I_{T,2} + I_{T,3} + I_{T,4}, \end{aligned}$$

where the definitions of $I_{T,j}$ for $j = 1, \ldots, 4$ are obvious.

By Lemma A.1.4, I have $I_{T,1} = O_P\left(\sqrt{d_T \log T}\right)$. Also, it's easy to see that $I_{T,2} = O_P(d_T)$ and $I_{T,3} = O_P(d_T)$, because $\|\widetilde{\mathbb{B}}_1(L)\varepsilon_0\| < \infty$ and $\|\widetilde{\mathbb{B}}_T(L)\varepsilon_T\| < \infty$ in view of the fact that

$$|\widetilde{\mathbb{B}}_1(1)| \leq \sum_{j=0}^{\infty} |\widetilde{\mathbf{B}}_{j,1}| < \infty \quad \text{and} \quad |\widetilde{\mathbb{B}}_T(1)| \leq \sum_{j=0}^{\infty} |\widetilde{\mathbf{B}}_{j,T}| < \infty$$

by Lemma A.1.3. Thus, I need only to consider $I_{T,4}$ below. Note that

(1). $\sum_{t=1}^{T-1} |\widetilde{\mathbb{B}}_{t+1}(1) - \widetilde{\mathbb{B}}_t(1)| = O(1)$ by Lemma A.1.3;

(2). $T^{2/\delta} d_T \log T \to 0$ and $\sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)| = O(d_T)$ by the conditions in the body of this lemma;

(3). $\max_{1\le t\le T-1} |\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t| = O_P(T^{1/\delta})$ by $\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t\|_\delta < \infty$ and

$$\max_{1\le t\le T-1} |\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t| \le \left(\sum_{t=1}^{T-1} |\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t|^\delta\right)^{1/\delta} = O_P(T^{1/\delta}).$$

Hence, write

$$\begin{aligned}
&\sup_{\tau\in[a,b]} \left|\sum_{t=1}^{T-1} \left(\mathbf{W}_{T,t+1}(\tau)\widetilde{\mathbb{B}}_{t+1}(L) - \mathbf{W}_{T,t}(\tau)\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\varepsilon}_t\right| \\
&= \sup_{\tau\in[a,b]} \left|\sum_{t=1}^{T-1} (\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t + \mathbf{W}_{T,t}(\tau)(\widetilde{\mathbb{B}}_{t+1}(L) - \widetilde{\mathbb{B}}_t(L))\boldsymbol{\varepsilon}_t\right| \\
&\le \max_{1\le t\le T-1} |\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\varepsilon}_t| \cdot \sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)| \\
&\quad + \sup_{\tau\in[a,b],1\le t\le T} |\mathbf{W}_{T,t}(\tau)| \cdot \sum_{t=1}^{T-1} |(\widetilde{\mathbb{B}}_{t+1}(L) - \widetilde{\mathbb{B}}_t(L))\boldsymbol{\varepsilon}_t| \\
&= O_P(T^{1/\delta} \cdot d_T) + O_P(d_T) = o_P(\sqrt{d_T \log T}).
\end{aligned}$$

The first result then follows.

(2). Below, I consider $p = 0$ only. The cases with fixed $p \ge 1$ can be verified in a similar manner, so omitted.

$$\begin{aligned}
&\sup_{\tau\in[a,b]} \left|\sum_{t=1}^{T} \text{vec}\left(\mathbf{W}_{T,t}(\tau)(\mathbf{x}_t\mathbf{x}_t^\top - E(\mathbf{x}_t\mathbf{x}_t^\top))\right)\right| \\
&\le 2 \sup_{\tau\in[a,b]} \left|\sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(\tau)) \sum_{j=0}^{\infty} (\mathbf{B}_{j,t} \otimes \boldsymbol{\mu}_t)\boldsymbol{\varepsilon}_{t-j}\right| \\
&+ \sup_{\tau\in[a,b]} \left|\sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(\tau)) \sum_{j=0}^{\infty} (\mathbf{B}_{j,t} \otimes \mathbf{B}_{j,t}) \left(\text{vec}\left(\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j}^\top\right) - \text{vec}(\mathbf{I}_d)\right)\right| \\
&+ 2 \sup_{\tau\in[a,b]} \left|\sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(\tau)) \sum_{r=1}^{\infty}\sum_{j=0}^{\infty} (\mathbf{B}_{j+r,t} \otimes \mathbf{B}_{j,t})\text{vec}(\varepsilon_{t-j}\varepsilon_{t-j-r})\right| := 2I_{T,1} + I_{T,2} + 2I_{T,3},
\end{aligned}$$

wherein $I_{T,1} = O_P(\sqrt{d_T \log T})$ by a proof similar to the first result of this lemma.

Consider $I_{T,2}$. Using Lemma A.1.3, write

$$
\begin{aligned}
I_{T,2} \le & \sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T}(\mathbf{I}_d\otimes \mathbf{W}_{T,t}(\tau))\mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)-\mathrm{vec}\left(\mathbf{I}_d\right)\right)\right| \\
& + \sup_{\tau\in[a,b]}\left|(\mathbf{I}_d\otimes \mathbf{W}_{T,1}(\tau))\widetilde{\mathbb{B}}_1^0(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_0\boldsymbol{\varepsilon}_0^\top\right)\right| + \sup_{\tau\in[a,b]}\left|(\mathbf{I}_d\otimes \mathbf{W}_{T,T}(\tau))\widetilde{\mathbb{B}}_T^0(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_T\boldsymbol{\varepsilon}_T^\top\right)\right| \\
& + \sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T-1}\left((\mathbf{I}_d\otimes \mathbf{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^0(L)-(\mathbf{I}_d\otimes \mathbf{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^0(L)\right)\cdot\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right| \\
:= & I_{T,21}+I_{T,22}+I_{T,23}+I_{T,24}.
\end{aligned}
$$

By Lemma A.1.5, I have $I_{T,21}=O_P\left(\sqrt{d_T\log T}\right)$. Also, $I_{T,22}=O_P(d_T)$ and $I_{T,23}=O_P(d_T)$, because $|\widetilde{\mathbb{B}}_1^0(1)|<\infty$ and $|\widetilde{\mathbb{B}}_T^0(1)|<\infty$ by Lemma A.1.3. Similar to the proof of the first result, for $I_{T,24}$, I write

$$
\begin{aligned}
& \sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T-1}\left((\mathbf{I}_d\otimes \mathbf{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^0(L)-(\mathbf{I}_d\otimes \mathbf{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^0(L)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right| \\
\le & \sqrt{d}\sup_{\tau\in[a,b],1\le t\le T}|\mathbf{W}_{T,t+1}(\tau)|\cdot\sum_{t=1}^{T-1}\left|\left(\widetilde{\mathbb{B}}_{t+1}^0(L)-\widetilde{\mathbb{B}}_t^0(L)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right| \\
& +\sqrt{d}\max_t\left|\widetilde{\mathbb{B}}_t^0(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right|\cdot\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}|\mathbf{W}_{T,t+1}(\tau)-\mathbf{W}_{T,t}(\tau)|=o_P\left(\sqrt{d_T\log T}\right),
\end{aligned}
$$

where I have used the following facts:

(1). $T^{4/\delta}d_T\log T\to 0$;

(2). $\max_{t\ge 1}\left|\widetilde{\mathbb{B}}_t^0(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)\right|=O_P(T^{2/\delta})$;

(3). $\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}|\mathbf{W}_{T,t+1}(\tau)-\mathbf{W}_{T,t}(\tau)|=O(d_T)$;

(4). $\sum_{t=1}^{T-1}\left|\widetilde{\mathbb{B}}_{t+1}^0(1)-\widetilde{\mathbb{B}}_t^0(1)\right|=O(1)$.

Then I can conclude that $I_{T,24}=O_P\left(\sqrt{d_T\log T}\right)$.

I now consider $I_{T,3}$. Using Lemma A.1.3, I have

$$
\begin{aligned}
& \sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T}(\mathbf{I}_d\otimes \mathbf{W}_{T,t}(\tau))\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\mathbf{B}_{j+r,t}\otimes\mathbf{B}_{j,t})\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j-r}\right)\right| \\
\le & \sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T}(\mathbf{I}_d\otimes \mathbf{W}_{T,t}(\tau))\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right| + \sup_{\tau\in[a,b]}\left|(\mathbf{I}_d\otimes \mathbf{W}_{T,1}(\tau))\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_1^r(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_0\boldsymbol{\varepsilon}_{-r}^\top\right)\right| \\
& + \sup_{\tau\in[a,b]}\left|(\mathbf{I}_d\otimes \mathbf{W}_{T,T}(\tau))\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_T^r(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_T\boldsymbol{\varepsilon}_{T-r}^\top\right)\right| \\
& + \sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left((\mathbf{I}_d\otimes \mathbf{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^r(L)-(\mathbf{I}_d\otimes \mathbf{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^r(L)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-r}^\top\right)\right| \\
:= & I_{T,31}+I_{T,32}+I_{T,33}+I_{T,34},
\end{aligned}
$$

where $\boldsymbol{\zeta}_t$ is defined in Lemma A.1.5.

By Lemma A.1.5, $I_{T,31} = O_P\left(\sqrt{d_T \log T}\right)$. Moreover, $I_{T,32} = O_P(d_T)$ and $I_{T,33} = O_P(d_T)$, because $\sum_{r=1}^{\infty} |\widetilde{\mathbb{B}}_1^r(1)| < \infty$ and $\sum_{r=1}^{\infty} |\widetilde{\mathbb{B}}_T^r(1)| < \infty$ by Lemma A.1.3. For $I_{T,34}$, I write

$$\begin{aligned}
&\sup_{\tau\in[a,b]} \left| \sum_{t=1}^{T-1}\sum_{r=1}^{\infty} \left((\mathbf{I}_d \otimes \mathbf{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^r(L) - (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^r(L) \right) \mathrm{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-r}^\top\right) \right| \\
\leq\ & \sqrt{d} \sup_{\tau\in[a,b],1\leq t\leq T} |\mathbf{W}_{T,t}(\tau)| \cdot \sum_{t=1}^{T-1} \left| \sum_{r=1}^{\infty} \left(\widetilde{\mathbb{B}}_{t+1}^r(L) - \widetilde{\mathbb{B}}_t^r(L) \right) \mathrm{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-r}^\top\right) \right| \\
&+\sqrt{d} \max_t \left| \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_t^r(L) \mathrm{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-r}^\top\right) \right| \cdot \sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)| = o_P(\sqrt{d_T \log T}),
\end{aligned}$$

where I have used the following results:

$$\begin{aligned}
&(1). \quad T^{4/\delta} d_T \log T \to 0; \\
&(2). \quad \max_t \left| \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_t^r(L) \mathrm{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-r}^\top\right) \right| = O_P(T^{2/\delta}); \\
&(3). \quad \sum_{t=1}^{T-1}\sum_{r=1}^{\infty} |\widetilde{\mathbb{B}}_{t+1}^r(1) - \widetilde{\mathbb{B}}_t^r(1)| = O(1); \\
&(4). \quad \sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)| = O(d_T).
\end{aligned}$$

Based on the above development, the proof of the case with $p = 0$ is done. The proof is now completed. □

Proof of Lemma 2.2.4.

Similar to the proof of Lemma 2.2.2, I have

$$\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \mathbf{W}_{T,t}(\mathbf{x}_t - E(\mathbf{x})) = \frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \mathbf{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t + o_P(1)$$

as $d_T = o(1)$.

Since

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \mathbf{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t \right) &= \frac{1}{d_T} \sum_{t=1}^{T} \mathbf{W}_{T,t}\mathbb{B}_t(1)\mathbb{B}_t^\top(1)\mathbf{W}_{T,t}^\top \\
&\to \boldsymbol{\Sigma}_{\mathbf{W}},
\end{aligned}$$

I then use the Cramér-Wold device to prove its asymptotic normality. That is to show that for any conformable vector $\mathbf{l}$,

$$\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \mathbf{l}^\top \mathbf{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t \to_D N\left(\mathbf{0}, \mathbf{l}^\top \boldsymbol{\Sigma}_{\mathbf{W}} \mathbf{l}\right).$$

Let $\mathbf{Z}_t = \frac{1}{\sqrt{d_T}} \mathbf{l}^\top \mathbf{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t$. By the law of large numbers for martingale differences and the assumption $E\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top | \mathscr{F}_{t-1}\right) = \mathbf{I}_d$ a.s., I have $\sum_{t=1}^{T} \mathbf{Z}_t^2(\tau) \to_P \mathbf{l}^\top \boldsymbol{\Sigma}_{\mathbf{W}} \mathbf{l}$.

Furthermore, for any $\nu > 0$, by both Hölder's and Markov's inequalities, I have

$$\begin{aligned}
&\sum_{t=1}^{T} E\left(\mathbf{Z}_t^2(\tau) I\left(|\mathbf{Z}_t(\tau)| > \nu\right)\right) \\
\leq &\sum_{t=1}^{T} \frac{1}{d_T} |\mathbf{W}_{T,t}|^\delta \left(E|\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t|^\delta\right)^{2/\delta} \left(\frac{E|\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t|^\delta}{(d_T)^{\delta/2}\nu^\delta}\right)^{(\delta-2)/\delta} \\
= &O\left(d_T^{(\delta-2)/2}\right) = o(1)
\end{aligned}$$

since $\sum_{t=1}^{T} |\mathbf{W}_{T,t}| = O(1)$ and $\max_{t\geq 1} |\mathbf{W}_{T,t}| = O(d_T)$. By Lemma A.1.1, the proof is now completed.

□

Proof of Lemma 2.2.5.

Let $\mathbf{e}_t = \mathbf{x}_t - E(\mathbf{x}_t)$ and $Z_T^* = \frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \mathbf{d}^\top \mathbf{W}_{T,t} \mathbf{e}_t \xi_t^*$ for any conformable unit vector $\mathbf{d}$. Then, it suffices to show that $Z_T^* \to_{D^*} N\left(\mathbf{0}, \mathbf{d}^\top \boldsymbol{\Sigma}_\mathbf{W} \mathbf{d}\right)$. In the following, I first show that

$$\mathrm{Var}^*(Z_T^*(\tau))^2 = \mathbf{d}^\top \boldsymbol{\Sigma}_\mathbf{W} \mathbf{d} + o_P(1)$$

and then prove its normality by blocking techniques.

Conditioning on the original sample, I have

$$\begin{aligned}
E^*(Z_T^*)^2 &= \frac{1}{d_T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{d}^\top \mathbf{W}_{T,t} \mathbf{e}_t \mathbf{e}_s^\top \mathbf{W}_{T,s}^\top \mathbf{d} E^*(\xi_t^* \xi_s^*) \\
= &\frac{1}{d_T} \sum_{t=1}^{T} \mathbf{d}^\top \mathbf{W}_{T,t} \mathbf{e}_t \mathbf{e}_t^\top \mathbf{W}_{T,t}^\top \mathbf{d} + \frac{1}{d_T} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \mathbf{d}^\top \mathbf{W}_{T,t} \mathbf{e}_t \mathbf{e}_{t+i}^\top \mathbf{W}_{T,t+i}^\top \mathbf{d} a(i/l) \\
&+ \frac{1}{d_T} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \mathbf{d}^\top \mathbf{W}_{T,t+i} \mathbf{e}_{t+i} \mathbf{e}_t^\top \mathbf{W}_{T,t}^\top \mathbf{d} a(i/l). \qquad \text{(A.2.2)}
\end{aligned}$$

For the first term on the right hand side of (A.2.2), similar to the proof of Lemma 2.2.2, it is straightforward to obtain that

$$\frac{1}{d_T} \sum_{t=1}^{T} \mathbf{d}^\top \mathbf{W}_{T,t} \mathbf{e}_t \mathbf{e}_t^\top \mathbf{W}_{T,t}^\top \mathbf{d} = \frac{1}{d_T} \sum_{t=1}^{T} \mathbf{d}^\top \mathbf{W}_{T,t} E(\mathbf{e}_t \mathbf{e}_t^\top) \mathbf{W}_{T,t}^\top \mathbf{d} + o_P(1).$$

For the second and third terms on the right hand side of (A.2.2), as $a(i/l) = 0$ for $i > l$, I have

$$\begin{aligned}
&E\left|\sum_{i=1}^{T-1} \frac{1}{d_T} \sum_{t=1}^{T-i} \mathbf{W}_{T,t} \left(\mathbf{e}_t \mathbf{e}_{t+i}^\top - E(\mathbf{e}_t \mathbf{e}_{t+i}^\top)\right) \mathbf{W}_{T,t+i}^\top a(i/l)\right| \\
\leq &\sum_{i=1}^{T-1} a(i/l) E\left|\frac{1}{d_T} \sum_{t=1}^{T-i} \mathbf{W}_{T,t} \left(\mathbf{e}_t \mathbf{e}_{t+i}^\top - E(\mathbf{e}_t \mathbf{e}_{t+i}^\top)\right) \mathbf{W}_{T,t+i}^\top\right| \\
= &l \cdot \sqrt{d_T} = o(1)
\end{aligned}$$

as I have $E\left|\frac{1}{d_T} \sum_{t=1}^{T-i} \mathbf{W}_{T,t} \left(\mathbf{e}_t \mathbf{e}_{t+i}^\top - E(\mathbf{e}_t \mathbf{e}_{t+i}^\top)\right) \mathbf{W}_{T,t+i}^\top\right| = O(\sqrt{d_T})$ by using similar arguments to those used in the proof of Lemma 2.2.2.

I now need only to focus on $\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\mathbf{W}_{T,t}E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\mathbf{W}_{T,t+i}^\top a(i/l)$. Note that

$$\begin{aligned}
&\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\mathbf{W}_{T,t}E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\mathbf{W}_{T,t+i}^\top a(i/l),\\
=\ &\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\mathbf{W}_{T,t}E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\mathbf{W}_{T,t+i}^\top\\
&+\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\mathbf{W}_{T,t}E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\mathbf{W}_{T,t+i}^\top(a(i/l)-1). \qquad \text{(A.2.3)}
\end{aligned}$$

It is then sufficient to show that the second term of the above equation is $o(1)$ since

$$\frac{1}{d_T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{W}_{T,t}E(\mathbf{e}_t\mathbf{e}_s^\top)\mathbf{W}_{T,s}^\top=\mathrm{Var}\left(\frac{1}{\sqrt{d_T}}\sum_{t=1}^{T}\mathbf{W}_{T,t}\mathbf{e}_t\right)\to\mathbf{\Sigma_W}$$

by the proof of Lemma 2.2.2.

Let s_T satisfy $\frac{1}{s_T}+\frac{s_T^2}{l}\to 0$. The second term of (A.2.3) is then bounded by

$$\begin{aligned}
&\left|\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\mathbf{W}_{T,t}E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\mathbf{W}_{T,t+i}^\top(a(i/l)-1)\right|\\
\leq &M\sum_{i=1}^{s_T}\max_t\left|E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\right||a(i/l)-1|+M\sum_{i=d_T+1}^{\infty}\max_t\left|E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\right||a(i/l)-1|\\
\leq &M\sum_{i=1}^{s_T}(1-a(i/l))+M\sum_{i=d_T+1}^{\infty}\max_t\left|E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\right|=o(1),
\end{aligned}$$

since $|\sum_{i=1}^{s_T}(1-a(i/l))|\leq M\sum_{i=1}^{s_T}i/l\leq Ms_T^2/l=o(1)$ by Lipschitz continuity of $a(\cdot)$ and

$$\sum_{i=s_T+1}^{\infty}\max_t\left|E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\right|=o(1)\text{ as }s_T\to\infty.$$

Conditioning on the original sample, I now use standard arguments for using a block technique to show the asymptotic normality. Now let $Z_T^*(\tau)=\sum_{j=1}^{k}X_{T,j}^*(\tau)+\sum_{j=1}^{k}Y_{T,j}^*(\tau)$, where

$$X_{T,j}^*(\tau)=\frac{1}{\sqrt{d_T}}\sum_{t=B_j+1}^{B_j+r_1}\mathbf{d}^\top\mathbf{W}_{T,t}\mathbf{e}_t\xi_t^*,\quad Y_{T,j}^*(\tau)=\frac{1}{\sqrt{d_T}}\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2}\mathbf{d}^\top\mathbf{W}_{T,t}\mathbf{e}_t\xi_t^*,$$

with $B_j=(j-1)(r_1+r_2)$ and $k=\lceil T/(r_1+r_2)\rceil$.

Let $r_1=r_1(T)$ and $r_2=r_2(T)$ satisfying $k\cdot r_2\cdot d_T\to 0$, $r_1\cdot d_T+l/(r_1)\to 0$ and $r_2/r_1+l/r_2\to 0$. I first show that $\sum_{j=1}^{k}Y_{T,j}^*(\tau)=o_P(1)$. Since $r_1>l$ for large enough T and the blocks $Y_{T,j}^*$ are mutual independent conditionally on the original data, then I have

$$
\begin{aligned}
EE^*\left(\sum_{j=1}^{k} Y^*_{T,j}(\tau)\right)^2 &= E\left(\sum_{j=1}^{k} E^*(Y^*_{T,j}(\tau))^2\right)\\
&\leq \frac{1}{d_T}\sum_{i=-r_2+1}^{r_2-1} a(i/l)\max_t \left|E(\mathbf{e}_t\mathbf{e}_{t+i}^\top)\right| \sum_{j=1}^{k}\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-i} |\mathbf{W}_{T,t}|\cdot|\mathbf{W}_{T,t+i}|\\
&\leq M\frac{1}{d_T}\max_{0\leq i\leq r_2-1}\sum_{j=1}^{k}\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-i} |\mathbf{W}_{T,t}|\cdot|\mathbf{W}_{T,t+i}| \leq \ Mkr_2d_T = o(1).
\end{aligned}
$$

I employ Lindeberg CLT to establish the asymptotic normality of $\sum_{j=1}^{k} X^*_{T,j}(\tau)$ as the blocks $X^*_{T,j}(\tau)$ are independent when $r_2 > l$ for large enough T. As discussed before, I have already shown that the asymptotic variance is equal to $\mathbf{\Sigma_W}$. I then need to verify that for every $\nu > 0$,

$$
\sum_{j=1}^{k} E^*\left(\frac{X^*_{T,j}(\tau)^2}{E^*\left(\sum_{j=1}^{k} X^*_{T,j}(\tau)\right)^2} I\left(\frac{X^*_{T,j}(\tau)^2}{E^*\left(\sum_{j=1}^{k} X^*_{T,j}(\tau)\right)^2} > \nu\right)\right) = o_P(1).
$$

Conditioning on the original sample, $\{\mathbf{e}_t\xi^*_t\}$ is an L_δ-mixingale sequence. By Hölder's inequality, Chebyshev's inequality and Lemma 2 in Hansen [1991], I have

$$
\begin{aligned}
&\sum_{j=1}^{k} E^*\left(\frac{X^*_{T,j}(\tau)^2}{E^*\left(\sum_{j=1}^{k} X^*_{T,j}(\tau)\right)^2} I\left(\frac{X^*_{T,j}(\tau)^2}{E^*\left(\sum_{j=1}^{k} X^*_{T,j}(\tau)\right)^2} > \nu\right)\right)\\
\leq\ & \sum_{j=1}^{k}\left(E^*\left(\frac{X^*_{T,j}(\tau)^2}{E^*(\sum_{j=1}^{k} X^*_{T,j}(\tau))^2}\right)^{\frac{\delta}{2}}\right)^{\frac{2}{\delta}}\left(\frac{E^*\left(\frac{X^*_{T,j}(\tau)^2}{E^*(\sum_{j=1}^{k} X^*_{T,j}(\tau))^2}\right)^{\frac{\delta}{2}}}{\nu^{\frac{\delta}{2}}}\right)^{\frac{\delta-2}{\delta}}\\
=\ & \nu^{\frac{2-\delta}{2}}\sum_{j=1}^{k}\frac{E^*(X^*_{T,j}(\tau))^\delta}{\left(E^*(\sum_{j=1}^{k} X^*_{T,j}(\tau))^2\right)^{\frac{\delta}{2}}} \leq \nu^{\frac{2-\delta}{2}}\sum_{j=1}^{k}\frac{d_T^{-\delta/2}M\left[\sum_{t=B_j+1}^{B_j+r_1}\left(\mathbf{d}^\top\mathbf{W}_{T,t}\mathbf{e}_t\right)^2\right]^{\delta/2}}{\left(E^*(\sum_{j=1}^{k} X^*_{T,j}(\tau))^2\right)^{\frac{\delta}{2}}}\\
\leq\ & \nu^{\frac{2-\delta}{2}}\sum_{j=1}^{k}\frac{Md_T^{-\delta/2}r_1^{\delta/2-1}\sum_{t=B_j+1}^{B_j+r_1}\left(\mathbf{d}^\top\mathbf{W}_{T,t}\mathbf{e}_t\right)^\delta}{\left(E^*(\sum_{j=1}^{k} X^*_{T,j}(\tau))^2\right)^{\frac{\delta}{2}}}\\
\leq\ & \nu^{\frac{2-\delta}{2}}\frac{Md_T^{-\delta/2}r_1^{\delta/2-1}d_T^{\delta-1}\sum_{t=1}^{T}|\mathbf{W}_{T,t}|\cdot|\mathbf{e}_t|^\delta}{\left(E^*(\sum_{j=1}^{k} X^*_{T,j}(\tau))^2\right)^{\frac{\delta}{2}}} = O_P((d_Tr_1)^{\delta/2-1}) = o_P(1).
\end{aligned}
$$

Combining the above results, I have

$$
Z^*_T \rightarrow_{D^*} N\left(\mathbf{0}, \mathbf{d}^\top\mathbf{\Sigma_W}\mathbf{d}\right).
$$

The proof is now completed. □

Proof of Lemma 2.2.6.

Define $\mathbf{\Xi}_i = \frac{1}{d_T}\sum_{t=1}^{T-i}\mathbf{W}_{T,t}E\left(\mathbf{e}_t\mathbf{e}_{t+i}^\top\right)\mathbf{W}_{T,t+i}^\top$ with $\mathbf{e}_t = \sum_{j=0}^{\infty}\mathbf{B}_{j,t}\boldsymbol{\varepsilon}_t$. Write

$$\begin{aligned}\widehat{\mathbf{\Sigma}}_{\mathbf{W}} &= \underbrace{\mathbf{\Xi}_0 + \sum_{i=1}^{b}\psi(i/b)\left(\mathbf{\Xi}_i + \mathbf{\Xi}_i^\top\right)}_{I_{T,1}} + \underbrace{\widehat{\mathbf{\Xi}}_0 - \mathbf{\Xi}_0}_{I_{T,2}} \\ &\quad + \underbrace{\sum_{i=1}^{b}\psi(i/b)\left(\widehat{\mathbf{\Xi}}_i - \mathbf{\Xi}_i + \widehat{\mathbf{\Xi}}_i^\top - \mathbf{\Xi}_i^\top\right)}_{I_{T,3}}.\end{aligned}$$

I next prove $\widehat{\mathbf{\Sigma}}_{\mathbf{W}} \to_P \mathbf{\Sigma}_{\mathbf{W}}$ by showing that $I_{T,1} \to \mathbf{\Sigma}_{\mathbf{W}}$, $I_{T,2} = o_P(1)$ and $I_{T,3} = o_P(1)$ one by one.

Consider $\sum_{i=1}^{b}\psi(i/b)\mathbf{\Xi}_i$. By the fact that $\max_t\sum_{j=1}^{\infty}j\,|\mathbf{B}_{j,t}| < \infty$, Lipschitz continuity of $\psi(\cdot)$, and $\psi(0)=1$, I have

$$\begin{aligned}\left|\sum_{i=1}^{b}(1-\psi(i/b))\mathbf{\Xi}_i\right| &\le M\cdot\max_t|\mathbf{W}_{T,t}|\cdot\frac{1}{d_T}\sum_{t=1}^{T}|\mathbf{W}_{T,t}|\sum_{j=0}^{\infty}|\mathbf{B}_{j,t}|\sum_{i=1}^{b}\frac{i}{b}|\mathbf{B}_{j+i,t+i}| \\ &= O(1/b) = o(1).\end{aligned}$$

Hence, I have $I_{T,1} \to \mathbf{\Sigma}_{\mathbf{W}}$.

For $I_{T,2}$ and $I_{T,3}$, since $\sum_{i=1}^{b}|\psi(i/b)| = O(b)$, $b\sqrt{d_T} \to 0$ and $E\|\widehat{\mathbf{\Xi}}_i - \mathbf{\Xi}_i\| = O(\sqrt{d_T})$ (using similar arguments to those used in the proofs of Lemma 2.2.2), I have

$$E\|I_{T,3}\| \le 2\max_{1\le i\le b}E\|\widehat{\mathbf{\Xi}}_i - \mathbf{\Xi}_i\|\cdot\sum_{i=1}^{b}|\psi(i/b)| = O(b\sqrt{d_T}) = o(1).$$

The proof is now completed. □

Proof of Theorem 2.3.1.

Since $\frac{1}{T}\sum_{t=1}^{T}K_h(\tau_t-\tau) = 1 + O\left(\frac{1}{Th}\right)$, I have

$$\widehat{\boldsymbol{\mu}}(\tau) - E\left(\widehat{\boldsymbol{\mu}}(\tau)\right) = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\mu}(\tau_t))K_h(\tau_t-\tau) + O_P\left(\frac{1}{Th}\right),$$

which follows that $\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}(\tau) - E\left(\widehat{\boldsymbol{\mu}}(\tau)\right)\right) = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\mu}(\tau_t))K\left(\frac{\tau_t-\tau}{h}\right) + o_P(1)$.

As $\frac{1}{T}\sum_{t=1}^{T}K_h(\tau_t-\tau) = O(1)$, $\max_t\frac{1}{T}K_h(\tau_t-\tau) = O(1/(Th))$ and $\frac{1}{T}\sum_{t=1}^{T-1}(K_h(\tau_{t+1}-\tau) - K_h(\tau_t-\tau)) = O(1/(Th))$, by Lemma 2.2.4, I have

$$\begin{aligned}&\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\mu}(\tau_t))K\left(\frac{\tau_t-\tau}{h}\right) \\ &= \frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\mathbb{B}_t(1)\boldsymbol{\varepsilon}_tK\left(\frac{\tau_t-\tau}{h}\right) + \frac{1}{\sqrt{Th}}\widetilde{\mathbb{B}}_1(L)\boldsymbol{\varepsilon}_0K\left(\frac{\tau_1-\tau}{h}\right) - \frac{1}{\sqrt{Th}}\widetilde{\mathbb{B}}_T(L)\boldsymbol{\varepsilon}_TK\left(\frac{\tau_T-\tau}{h}\right) \\ &\quad + \frac{1}{\sqrt{Th}}\sum_{t=1}^{T-1}\left(\widetilde{\mathbb{B}}_{t+1}(L)K\left(\frac{\tau_{t+1}-\tau}{h}\right) - \widetilde{\mathbb{B}}_t(L)K\left(\frac{\tau_t-\tau}{h}\right)\right)\boldsymbol{\varepsilon}_t \\ &\to_D N\left(\mathbf{0}, \widetilde{v}_0\left\{\sum_{j=0}^{\infty}\mathbf{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty}\mathbf{B}_j^\top(\tau)\right\}\right).\end{aligned}$$

For the bias term, I have for any $\tau \in (0,1)$

$$\frac{1}{Th}\sum_{t=1}^{T}\boldsymbol{\mu}(\tau_t)K\left(\frac{\tau_t-\tau}{h}\right)=\boldsymbol{\mu}(\tau)+\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\mu}^{(2)}(\tau)+o(h^2)+O\left(\frac{1}{Th}\right).$$

The proof is now completed. □

Proof of Theorem 2.3.2.

Note that $\mathbf{x}_t^*=\widetilde{\boldsymbol{\mu}}(\tau_t)+\mathbf{e}_t^*$, so I can write

$$\widehat{\boldsymbol{\mu}}^*(\tau)-\widetilde{\boldsymbol{\mu}}(\tau)=\left(\sum_{t=1}^{T}W_{T,t}(\tau)\widetilde{\boldsymbol{\mu}}(\tau_t)-\widetilde{\boldsymbol{\mu}}(\tau)\right)+\sum_{t=1}^{T}W_{T,t}(\tau)\mathbf{e}_t^*:=\mathbf{I}_{T,1}+\mathbf{I}_{T,2},$$

where $W_{T,t}(\tau)=K(\frac{\tau_t-\tau}{h})/\sum_{t=1}^{T}K(\frac{\tau_t-\tau}{h})$.

I start our investigation from $\mathbf{I}_{T,1}$, and write

$$\begin{aligned}\mathbf{I}_{T,1} &= \left(\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\frac{1}{T\widetilde{h}}\sum_{s=1}^{T}\boldsymbol{\mu}(\tau_s)K\left(\frac{\tau_s-\tau_t}{\widetilde{h}}\right)-\frac{1}{T\widetilde{h}}\sum_{s=1}^{T}\boldsymbol{\mu}(\tau_s)K\left(\frac{\tau_s-\tau}{\widetilde{h}}\right)\right)\\ &\quad+\frac{1}{\sqrt{T\widetilde{h}}}\left(\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\mathbf{Z}_T(\tau_t)-\mathbf{Z}_T(\tau)\right)+O_P\left(\frac{1}{Th}\right)\\ &:= \mathbf{I}_{T,11}+\mathbf{I}_{T,12}+O_P\left(\frac{1}{Th}\right),\end{aligned}$$

where the definitions of $\mathbf{I}_{T,11}$ and $\mathbf{I}_{T,12}$ should be obvious, $\mathbf{Z}_T(\tau)=\frac{1}{\sqrt{T\widetilde{h}}}\sum_{t=1}^{T}\mathbf{e}_tK\left(\frac{\tau_t-\tau}{\widetilde{h}}\right)$ and $\mathbf{e}_t=\sum_{j=0}^{\infty}\mathbf{B}_j(\tau_t)\boldsymbol{\varepsilon}_{t-j}$. Similar to the development of Lemma 2.2.2, I can show that $\|\mathbf{I}_{T,12}\|=O_P((T\widetilde{h})^{-1/2})$, which, along with the conditions of Theorem 2.3.2, yield

$$\sqrt{Th}|\mathbf{I}_{T,12}|=O_P((h/\widetilde{h})^{1/2})=o_P(1).$$

For $\mathbf{I}_{T,11}$, by the definition of Riemann integral, I have

$$\begin{aligned}\mathbf{I}_{T,11} &= \int_{-1}^{1}K(u)\int_{-1}^{1}K(v)\left(\boldsymbol{\mu}(\tau+v\widetilde{h}+uh)-\boldsymbol{\mu}(\tau+v\widetilde{h})\right)\mathrm{d}v\mathrm{d}u+O\left(\frac{1}{Th}\right)\\ &=\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\mu}^{(2)}(\tau)+O(h^2(h+\widetilde{h}))+O\left(\frac{1}{Th}\right).\end{aligned}$$

Thus, I need only to focus on $\mathbf{I}_{T,2}$ and then show that

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\mathbf{e}_t^*\to_{D^*}N\left(\mathbf{0},\widetilde{v}_0\left\{\sum_{j=0}^{\infty}\mathbf{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty}\mathbf{B}_j^\top(\tau)\right\}\right).$$

Using the Cramér-Wold device, this is enough to show for any conformable unit vector $\mathbf{d}$,

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\mathbf{d}^\top\mathbf{e}_t^*\to_{D^*}N\left(\mathbf{0},\widetilde{v}_0\mathbf{d}^\top\left\{\sum_{j=0}^{\infty}\mathbf{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty}\mathbf{B}_j^\top(\tau)\right\}\mathbf{d}\right).$$

For $\forall \tau \in [h+\widetilde{h}, 1-h-\widetilde{h}]$, I write

$$\begin{aligned}
&\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\mathbf{d}^\top \widetilde{\mathbf{e}}_t \xi_t^* \\
=\;& Z_T^*(\tau) + \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\mathbf{d}^\top (\widetilde{\mathbf{e}}_t - \mathbf{e}_t)\xi_t^* \\
=\;& Z_T^*(\tau) + o_{P^*}(1), \qquad\qquad (A.2.4)
\end{aligned}$$

where $Z_T^*(\tau) = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\mathbf{d}^\top \mathbf{e}_t\xi_t^*$, and the second equality follows from

$$\begin{aligned}
& EE^*\left|\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\mathbf{d}^\top (\widetilde{\mathbf{e}}_t - \mathbf{e}_t)\xi_t^*\right|^2 \\
\le & \max_{\lfloor T(\tau-h)\rfloor \le t \le \lceil T(\tau+h)\rceil} E\left|\widetilde{\mathbf{e}}_t - \mathbf{e}_t\right|^2 \left(\frac{1}{Th}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right) K\left(\frac{\tau_s-\tau}{h}\right) E^*(\xi_t^*\xi_s^*)\right) \\
= & O\left(\widetilde{h}^4 + 1/(T\widetilde{h})\right) O(l) = o(1),
\end{aligned}$$

where $EE^*[\cdot]$ stands for taking the expectation of the variables with respect to the bootstrap draws, and then taking the exception with respect to the original sample.

By Lemma 2.2.5, I already have

$$Z_T^*(\tau) \to_{D^*} N\left(0, \widetilde{v}_0 \mathbf{d}^\top \left\{\sum_{j=0}^{\infty} \mathbf{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty} \mathbf{B}_j^\top(\tau)\right\}\mathbf{d}\right).$$

Combining the above results, I have

$$\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)\right) \to_{D^*} N\left(\boldsymbol{\mu}_b(\tau), \widetilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)\right). \qquad (A.2.5)$$

The proof is now completed.

□

Proof of Theorem 2.3.3.

Define $\boldsymbol{\Xi}_i(\tau) = \frac{1}{Th}\sum_{t=1}^{T-i} E\left(\mathbf{e}_t\mathbf{e}_{t+i}^\top\right) K\left(\frac{\tau_t-\tau}{h}\right)$ with $\mathbf{e}_t = \sum_{j=0}^{\infty}\mathbf{B}_j(\tau_t)\boldsymbol{\varepsilon}_t$. Write

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) =\;& \underbrace{\boldsymbol{\Xi}_0(\tau) + \sum_{i=1}^{b}\psi(i/b)\left(\boldsymbol{\Xi}_i(\tau) + \boldsymbol{\Xi}_i^\top(\tau)\right)}_{I_{T,1}} + \underbrace{\widehat{\boldsymbol{\Xi}}_0(\tau) - \boldsymbol{\Xi}_0(\tau)}_{I_{T,2}} \\
& + \underbrace{\sum_{i=1}^{b}\psi(i/b)\left(\widehat{\boldsymbol{\Xi}}_i(\tau) - \boldsymbol{\Xi}_i(\tau) + \widehat{\boldsymbol{\Xi}}_i^\top(\tau) - \boldsymbol{\Xi}_i^\top(\tau)\right)}_{I_{T,3}}.
\end{aligned}$$

I next prove $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) \to_P \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$ by showing that $I_{T,1} \to \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$, $I_{T,2} = o_P(1)$ and $I_{T,3} = o_P(1)$ one by one.

First, consider $I_{T,1}$. For $\boldsymbol{\Xi}_0(\tau)$, since $\sum_{j=0}^{\infty}\left|\mathbf{B}_j(\tau)\mathbf{B}_j^\top(\tau)\right| \le \left(\sum_{j=0}^{\infty}|\mathbf{B}_j(\tau)|\right)^2 < \infty$, $\mathbb{B}_0(\tau) := \sum_{j=0}^{\infty}\mathbf{B}_j(\tau)\mathbf{B}_j^\top(\tau)$ converges uniformly over $[0,1]$ and is second–order continuously differentiable with $\mathbb{B}^{(1)}(\tau) = \sum_{j=0}^{\infty} \left(\mathbf{B}_j^{(1)}(\tau)\mathbf{B}_j^\top(\tau) + \mathbf{B}_j(\tau)\mathbf{B}_j^{(1),\top}(\tau) \right)$ and

$$\mathbb{B}^{(2)}(\tau) = \sum_{j=0}^{\infty} \left(\mathbf{B}_j^{(2)}(\tau)\mathbf{B}_j^\top(\tau) + 2\mathbf{B}_j^{(1)}(\tau)\mathbf{B}_j^{(1),\top}(\tau) + \mathbf{B}_j(\tau)\mathbf{B}_j^{(2),\top}(\tau) \right).$$

Hence,

$$\begin{aligned}
\mathbf{\Xi}_0(\tau) &= \frac{1}{Th}\sum_{t=1}^{T}\sum_{j=0}^{\infty} \mathbf{B}_j(\tau_t)\mathbf{B}_j^\top(\tau_t) K\left(\frac{\tau_t - \tau}{h}\right) \\
&= \int_{-1}^{1} \mathbb{B}(\tau + uh) K(u)\,\mathrm{d}u + O(1/(Th)) \\
&= \sum_{j=0}^{\infty} \mathbf{B}_j(\tau)\mathbf{B}_j^\top(\tau) + o(1).
\end{aligned}$$

Consider $\sum_{i=1}^{b} \psi(i/b)\,\mathbf{\Xi}_i(\tau)$. By the fact that $\sup_{\tau\in[0,1]}\sum_{j=1}^{\infty} j\,|\mathbf{B}_j(\tau)| < \infty$, the Lipschitz continuity of $\psi(\cdot)$, and $\psi(0) = 1$, I have

$$\begin{aligned}
\left|\sum_{i=1}^{b} (1 - \psi(i/b))\,\mathbf{\Xi}_i(\tau)\right| &\leq M \cdot \left(\frac{1}{Th}\sum_{t=1}^{T}\sum_{j=0}^{\infty} |\mathbf{B}_j(\tau_t)| \sum_{i=1}^{b} \frac{i}{b} |\mathbf{B}_{j+i}(\tau_{t+i})| K\left(\frac{\tau_t - \tau}{h}\right) \right) \\
&= O(1/b) = o(1).
\end{aligned}$$

Hence, $\sum_{i=1}^{b} \psi(i/b)\,\mathbf{\Xi}_i(\tau) = \sum_{i=1}^{b} \mathbf{\Xi}_i(\tau) + o(1)$.

Since

$$\sum_{i=1}^{b}\sum_{j=0}^{\infty} \left|\mathbf{B}_j(\tau)\mathbf{B}_{j+i}^\top(\tau)\right| \leq \left(\sum_{j=0}^{\infty} |\mathbf{B}_j(\tau)|\right)^2 < \infty,$$

$\mathbb{B}_b(\tau) := \sum_{i=1}^{b}\sum_{j=0}^{\infty} \mathbf{B}_j(\tau)\mathbf{B}_{j+i}^\top(\tau)$ converges uniformly over $[0,1]$ and is second–order continuously differentiable with $\mathbb{B}_b^{(1)}(\tau) = \sum_{i=1}^{b}\sum_{j=0}^{\infty} \left(\mathbf{B}_j^{(1)}(\tau)\mathbf{B}_{j+i}^\top(\tau) + \mathbf{B}_j(\tau)\mathbf{B}_{j+i}^{(1),\top}(\tau) \right)$ and

$$\mathbb{B}_b^{(2)}(\tau) = \sum_{i=1}^{b}\sum_{j=0}^{\infty} \left(\mathbf{B}_j^{(2)}(\tau)\mathbf{B}_{j+i}^\top(\tau) + 2\mathbf{B}_j^{(1)}(\tau)\mathbf{B}_{j+i}^{(1),\top}(\tau) + \mathbf{B}_j(\tau)\mathbf{B}_j^{(2),\top}(\tau) \right).$$

In addition, since

$$\begin{aligned}
\left|\mathbb{B}_b(\tau_t) - \sum_{i=1}^{b}\sum_{j=0}^{\infty} \mathbf{B}_j(\tau_t)\mathbf{B}_{j+i}^\top(\tau_{t+i})\right| &\leq \sum_{i=1}^{b}\sum_{j=0}^{\infty} |\mathbf{B}_j(\tau_t)| \cdot |\mathbf{B}_{j+i}(\tau_t) - \mathbf{B}_{j+i}(\tau_{t+i})| \\
&\leq M \sum_{j=0}^{\infty} |\mathbf{B}_j(\tau_t)| \cdot \sum_{i=1}^{\infty} \left| \frac{i}{T} \mathbf{B}_i^{(1)}(\tau_t) \right| = O(1/T),
\end{aligned}$$

I have

$$\begin{aligned}
\sum_{i=1}^{b} \mathbf{\Xi}_i(\tau) &= \sum_{i=1}^{b} \frac{1}{Th} \sum_{t=1}^{T-i}\sum_{j=0}^{\infty} \mathbf{B}_j(\tau_t)\mathbf{B}_{j+i}^\top(\tau_{t+i}) K\left(\frac{\tau_t - \tau}{h}\right) \\
&= \int_{-1}^{1} \mathbb{B}_b(\tau + uh) K(u)\,\mathrm{d}u + O(1/(Th)) = \sum_{i=1}^{b}\sum_{j=0}^{\infty} \mathbf{B}_j(\tau)\mathbf{B}_{j+i}^\top(\tau) + o(1).
\end{aligned}$$

Hence, as $b \to \infty$, $I_{T,1} \to \mathbf{\Sigma}_{\boldsymbol{\mu}}(\tau)$.

For $I_{T,2}$, by Lemma 2.2.2 and $\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right) = 1 + O\left(1/(Th)\right)$, I have $I_{T,2} = o_P(1)$. Next, consider $I_{T,3}$. Since $\sum_{i=1}^{b} |\psi(i/b)| = O(b)$ and $b\max\{h^4, 1/\sqrt{Th}\} \to 0$, I have

$$E\left|I_{T,3}\right| \leq 2\max_{1\leq i\leq b} E\left|\widehat{\mathbf{\Xi}}_i(\tau) - \mathbf{\Xi}_i(\tau)\right| \cdot \sum_{i=1}^{b} |\psi(i/b)| = O(h^4 + 1/\sqrt{Th}) \cdot b = o(1)$$

if

$$\max_{1\leq i\leq b} E\left|\widehat{\mathbf{\Xi}}_i(\tau) - \mathbf{\Xi}_i(\tau)\right| = O\left(h^4 + 1/\sqrt{Th}\right). \tag{A.2.6}$$

Thus, I need only to complete the proof by proving (A.2.6). Thus, I write

$$\begin{aligned}
&\max_{1\leq i\leq b} E\left|\widehat{\mathbf{\Xi}}_i(\tau) - \mathbf{\Xi}_i(\tau)\right| \\
\leq\ & \max_{1\leq i\leq b} E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\mathbf{e}_t\mathbf{e}_{t+i}^\top - E\left(\mathbf{e}_t\mathbf{e}_{t+i}^\top\right)\right) K\left(\frac{\tau_t-\tau}{h}\right)\right| \\
& + \max_{1\leq i\leq b} E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\mathbf{e}}_t\widehat{\mathbf{e}}_{t+i}^\top - \mathbf{e}_t\mathbf{e}_{t+i}^\top\right) K\left(\frac{\tau_t-\tau}{h}\right)\right| + O(1/(Th)) \\
=\ & I_{T,4} + I_{T,5} + O(1/(Th)),
\end{aligned}$$

where the definitions of $I_{T,4}$ and $I_{T,5}$ should be obvious. By Lemma 2.2.2, I have $I_{T,4} = O(1/\sqrt{Th})$.

For $I_{T,5}$, write

$$\begin{aligned}
& E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\mathbf{e}}_t\widehat{\mathbf{e}}_{t+i}^\top - \mathbf{e}_t\mathbf{e}_{t+i}^\top\right) K\left(\frac{\tau_t-\tau}{h}\right)\right| \\
=\ & E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\mathbf{e}}_t - \mathbf{e}_t\right)\left(\widehat{\mathbf{e}}_{t+i} - \mathbf{e}_{t+i}\right)^\top K\left(\frac{\tau_t-\tau}{h}\right)\right| + E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\mathbf{e}}_t - \mathbf{e}_t\right)\mathbf{e}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)\right| \\
& + E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\mathbf{e}_t\left(\widehat{\mathbf{e}}_{t+i} - \mathbf{e}_{t+i}\right)^\top K\left(\frac{\tau_t-\tau}{h}\right)\right| = I_{T,51} + I_{T,52} + I_{T,53}.
\end{aligned}$$

For $I_{T,51}$, by Cauchy–Schwarz inequality and Theorem 2.3.1,

$$I_{T,51} \leq \max_{\lceil T(\tau-h)\rceil\leq t\leq\lceil T(\tau+h)\rceil} E\left|\widehat{\mathbf{e}}_t - \mathbf{e}_t\right|^2 \frac{1}{Th}\sum_{t=1}^{T-i} K\left(\frac{\tau_t-\tau}{h}\right) = O(h^4 + 1/(Th)).$$

For $I_{T,52}$, again using $\frac{1}{Th}\sum_{t=1}^{T-i} K\left(\frac{\tau_t-\tau}{h}\right) = 1 + O\left(1/(Th)\right)$, I have

$$\begin{aligned}
I_{T,52} \leq\ & E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\mathbf{M}_{\boldsymbol{\mu}}(\tau_t)\mathbf{e}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)\right| \\
& + E\left|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\frac{1}{Th}\sum_{s=1}^{T}\mathbf{e}_s K\left(\frac{\tau_t-\tau_s}{h}\right)\right)\mathbf{e}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)\right| + O(1/(Th)) \\
=\ & I_{T,521} + I_{T,522},
\end{aligned}$$

where $\mathbf{M}_{\boldsymbol{\mu}}(\tau) = \boldsymbol{\mu}(\tau) - \frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{\mu}(\tau_s)K\left(\frac{\tau_s-\tau}{h}\right)$ is a twice-differential function matrix satisfying that $\mathbf{M}_{\boldsymbol{\mu}}(\tau) = O(h^2)$. Hence, by Lemma 2.2.2, I have $I_{T,521} = O(h^2/\sqrt{Th})$. For $I_{T,522}$, by

Cauchy–Schwarz inequality, I have

$$\begin{aligned} I_{T,522} &\leq \frac{1}{Th}\sum_{t=1}^{T-i}\left\{E\left|\frac{1}{Th}\sum_{s=1}^{T}\mathbf{e}_sK\left(\frac{\tau_t-\tau_s}{h}\right)\right|^2\right\}^{1/2}\{E\,|\mathbf{e}_{t+i}|^2\}^{1/2}K\left(\frac{\tau_t-\tau}{h}\right) \\ &\leq \max_t\left\{E\left|\frac{1}{Th}\sum_{s=1}^{T}\mathbf{e}_sK\left(\frac{\tau_t-\tau_s}{h}\right)\right|^2\right\}^{1/2}\frac{1}{Th}\sum_{t=1}^{T-i}\{E\,|\mathbf{e}_{t+i}|^2\}^{1/2}K\left(\frac{\tau_t-\tau}{h}\right) \\ &= O(1/\sqrt{Th}). \end{aligned}$$

Putting the above results together, the proof is now completed. □

Proof of Theorem 2.4.1.

(1). For notational simplicity, let $\mathbf{Z}_{\tau,t}$ be the transpose of the t^{th} row of $\mathbf{Z}_\tau$. Also, I define

$$\begin{aligned} \mathbf{S}_{T,k}(\tau) &= \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{Z}_t^\top\left(\frac{\tau_t-\tau}{h}\right)^kK_h(\tau_t-\tau) \text{ for } 0\leq k\leq 3, \\ \mathbf{M}(\tau_t) &= \boldsymbol{\beta}(\tau_t)-\boldsymbol{\beta}(\tau)-\boldsymbol{\beta}^{(1)}(\tau)(\tau_t-\tau)-\frac{1}{2}\boldsymbol{\beta}^{(2)}(\tau)(\tau_t-\tau)^2, \\ \mathbf{S}_T(\tau) &= \begin{pmatrix}\mathbf{S}_{T,0}(\tau) & \mathbf{S}_{T,1}(\tau)\\ \mathbf{S}_{T,1}(\tau) & \mathbf{S}_{T,2}(\tau)\end{pmatrix}. \end{aligned}$$

Since

$$\mathbf{y}_t=\mathbf{Z}_t^\top\left(\boldsymbol{\beta}(\tau)+\boldsymbol{\beta}^{(1)}(\tau)(\tau_t-\tau)+\frac{1}{2}\boldsymbol{\beta}^{(2)}(\tau)(\tau_t-\tau)^2+\mathbf{M}(\tau_t)\right)+\boldsymbol{\eta}_t,$$

I can write

$$\begin{aligned} &\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau) \\ =\ &[\mathbf{I}_l,\mathbf{0}_l]\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{\tau,t}\mathbf{Z}_{,\tau,t}^\top K_h(\tau_t-\tau)\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{\tau,t}\mathbf{y}_tK_h(\tau_t-\tau)-\boldsymbol{\beta}(\tau) \\ =\ &[\mathbf{I}_l,\mathbf{0}_l]\mathbf{S}_T^{-1}(\tau)\begin{bmatrix}\mathbf{S}_{T,2}(\tau)\\ \mathbf{S}_{T,3}(\tau)\end{bmatrix}\frac{1}{2}h^2\boldsymbol{\beta}^{(2)}(\tau)+[\mathbf{I}_l,\mathbf{0}_l]\mathbf{S}_T^{-1}(\tau)\cdot\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{\tau,t}\mathbf{Z}_t^\top\mathbf{M}(\tau_t)K_h(\tau_t-\tau) \\ &+[\mathbf{I}_l,\mathbf{0}_l]\mathbf{S}_T^{-1}(\tau)\cdot\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{\tau,t}\boldsymbol{\eta}_tK_h(\tau_t-\tau)=I_{T,1}+I_{T,2}+I_{T,3}, \end{aligned}$$

where the definitions of $I_{T,1}$ to $I_{T,3}$ should be obvious.

By Lemma A.1.6 and Lemma 2.2.2, I have

$$\begin{aligned} I_{T,1} &= \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\beta}^{(2)}(\tau)+O_P\left(h^2(h^2+(Th)^{-1/2})\right), \\ I_{T,2} &= o_P(h^2). \end{aligned}$$

Thus, I focus on $I_{T,3}$ below. For any $\tau\in(0,1)$, as $\{\mathbf{Z}_t\boldsymbol{\eta}_t\}$ is a sequence of martingale differences, by Lemma 2.2.2 and the martingale central limit theory, I have

$$\sqrt{Th}I_{T,3}=\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\tau)\otimes\mathbf{I}_d\right)\left(\frac{\sqrt{Th}}{T}\sum_{t=1}^{T}\mathbf{Z}_t\boldsymbol{\eta}_tK_h(\tau_t-\tau)\right)+o_P(1)\to_D N\left(\mathbf{0},\widetilde{v}_0\mathbf{V}(\tau)\right).$$

The proof of the first result of this theorem is now completed.

(2). The uniform convergence rate for $\widehat{\boldsymbol{\beta}}(\tau)$ follows directly from Lemmas 2.2.3 and A.1.7.3.

(3). By Lemma 2.2.2, I have

$$\left|\widehat{\boldsymbol{\Sigma}}_{\mathbf{z}}(\tau) - \boldsymbol{\Sigma}_{\mathbf{z}}(\tau)\right| = o_P(1).$$

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

Write

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

Consider $I_{T,1}$. Since $\{\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top - E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top)\}$ is a sequence of martingale differences, I have

$$\left|\frac{1}{T}\sum_{t=1}^T \left[\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top - E\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top\right)\right] K_h(\tau_t-\tau)\right| = o_P(1).$$

Next, consider $I_{T,2}$. By the second result of this theorem

$$|I_{T,2}| \leq \sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)|^2 \cdot \frac{1}{T}\sum_{t=1}^T |\mathbf{Z}_t|^2 K_h(\tau_t-\tau) = o_P(1).$$

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

$$|I_{T,3}| \leq \sup_{\tau\in[0,1]} \left|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right| \cdot \frac{1}{T}\sum_{t=1}^T |\mathbf{Z}_t\boldsymbol{\eta}_t| K_h(\tau_t-\tau) = o_P(1).$$

The proof is now completed. □

Proof of Theorem 2.4.2.

(1). By Lemma A.1.8,

$$\begin{aligned}
&\sqrt{T}(\widehat{\mathbf{c}} - \mathbf{c}) \\
&= \left(\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT} - \mathbf{S})^\top(\mathbf{I}_{dT} - \mathbf{S})\mathbf{X}_{\mathbf{C}}\right)^{-1}\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT} - \mathbf{S})^\top(\mathbf{I}_{dT} - \mathbf{S})\left(\begin{bmatrix}\mathbf{X}_{\widetilde{\mathbf{C}},1}^\top\boldsymbol{\theta}(\tau_1) \\ \vdots \\ \mathbf{X}_{\widetilde{\mathbf{C}},T}^\top\boldsymbol{\theta}(\tau_T)\end{bmatrix} + \boldsymbol{\eta}\right) \\
&= \boldsymbol{\Sigma}^{-1}\frac{1}{\sqrt{T}}\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT} - \mathbf{S})^\top(\mathbf{I}_{dT} - \mathbf{S})\boldsymbol{\eta} + o_P(1).
\end{aligned}$$

Hence, it suffices to show that

$$\frac{1}{\sqrt{T}}\mathbf{X}_{\mathbf{C}}^{\top}(\mathbf{I}_{dT}-\mathbf{S})^{\top}(\mathbf{I}_{dT}-\mathbf{S})\boldsymbol{\eta} \to_D N\left(\mathbf{0},\boldsymbol{\Delta}\right).$$

By using the same argument as in the proof of Lemma A.1.8,

$$\frac{1}{\sqrt{T}}\mathbf{X}_{\mathbf{C}}^{\top}(\mathbf{I}_{dT}-\mathbf{S})^{\top}(\mathbf{I}_{dT}-\mathbf{S})\boldsymbol{\eta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{X}_{\mathbf{C},t}-\boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau_t)\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}(\tau_t)\mathbf{X}_{\widetilde{\mathbf{C}},t}\right)\boldsymbol{\eta}_t + o_P(1).$$

Since $\left\{\left(\mathbf{X}_{\mathbf{C},t}-\boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau_t)\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}(\tau_t)\mathbf{X}_{\widetilde{\mathbf{C}},t}\right)\boldsymbol{\eta}_t\right\}$ is a sequence of martingale differences, the result follows by the central limit theorem for martingale differences. Note that the convergence of conditional variance can be proved by Lemma 2.2.2.

(2). Let $\boldsymbol{\Theta}(\tau)=[\boldsymbol{\theta}(\tau)^{\top}, h\boldsymbol{\theta}^{(1)}(\tau)^{\top}]^{\top}$. Note that

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau) &= [\mathbf{I}_{l-s},\mathbf{0}_{l-s}](\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\mathbf{X}_{\widetilde{\mathbf{C}},\tau})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\left(\mathbf{y}-\mathbf{X}_{\mathbf{C}}\widehat{\mathbf{c}}-\mathbf{X}_{\widetilde{\mathbf{C}},\tau}\boldsymbol{\Theta}(\tau)\right)\\
&= [\mathbf{I}_{l-s},\mathbf{0}_{l-s}](\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\mathbf{X}_{\widetilde{\mathbf{C}},\tau})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\mathbf{X}_{\mathbf{C}}(\mathbf{c}-\widehat{\mathbf{c}})\\
&\quad +[\mathbf{I}_{l-s},\mathbf{0}_{l-s}](\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\mathbf{X}_{\widetilde{\mathbf{C}},\tau})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\left(\begin{bmatrix}\mathbf{X}_{\widetilde{\mathbf{C}},1}^{\top}\boldsymbol{\theta}(\tau_1)\\ \vdots \\ \mathbf{X}_{\widetilde{\mathbf{C}},T}^{\top}\boldsymbol{\theta}(\tau_T)\end{bmatrix}-\mathbf{X}_{\widetilde{\mathbf{C}},\tau}\boldsymbol{\Theta}(\tau)\right)\\
&\quad +[\mathbf{I}_{l-s},\mathbf{0}_{l-s}](\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\mathbf{X}_{\widetilde{\mathbf{C}},\tau})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\boldsymbol{\eta}\\
&:= \mathbf{I}_{T,1}+\mathbf{I}_{T,2}+\mathbf{I}_{T,3}.
\end{aligned}$$

By part (1), I have $\mathbf{I}_{T,1}=O_P(T^{-1/2})$. By using standard arguments of the local linear kernel method, I have $\mathbf{I}_{T,2}=\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\theta}^{(2)}(\tau)+o_P(h^2)$. Then, it suffices to show that

$$[\mathbf{I}_{l-s},\mathbf{0}_{l-s}]\frac{\sqrt{h}}{\sqrt{T}}\mathbf{X}_{\widetilde{\mathbf{C}},\tau}^{\top}\mathbf{K}_{\tau}\boldsymbol{\eta}\to_D N\left(\mathbf{0},\tilde{v}_0\widetilde{\mathbf{C}}\left(\boldsymbol{\Sigma}_{\mathbf{z}}(\tau)\otimes\boldsymbol{\Omega}(\tau)\right)\widetilde{\mathbf{C}}^{\top}\right).$$

The above result follows by the central limit theorem for martingale differences.

□

A.3 Proofs of the Preliminary Lemmas

This section gives the proofs of the preliminary lemmas of Section A.1.

Proof of Lemma A.1.3.

(1). The first result follows from the standard BN decomposition (e.g., Phillips and Solo, 1992), so the details are omitted.

(2). For the second decomposition, write

$$
\begin{aligned}
& (1-L)\widetilde{\mathbb{B}}_t^r(L) \\
= \; & \sum_{j=0}^{\infty}\left(L^j \sum_{k=j+1}^{\infty}(\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t}) - L^{j+1}\sum_{k=j+1}^{\infty}(\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t})\right) \\
= \; & \sum_{j=0}^{\infty}\left(L^{j+1} \sum_{k=j+2}^{\infty}(\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t}) - L^{j+1}\sum_{k=j+1}^{\infty}(\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t})\right) + \sum_{k=1}^{\infty}(\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t}) \\
= \; & -\sum_{j=0}^{\infty}L^{j+1}(\mathbf{B}_{j+1+r,t}\otimes\mathbf{B}_{j+1,t}) + \sum_{k=1}^{\infty}(\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t}) \\
= \; & -\sum_{j=0}^{\infty}L^{j}(\mathbf{B}_{j+r,t}\otimes\mathbf{B}_{j,t}) + \sum_{k=0}^{\infty}(\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t}) = \mathbb{B}_t^r(1) - \mathbb{B}_t^r(L).
\end{aligned}
$$

(3). By Assumption 2.2.1,

$$
\max_{t\ge 1}\sum_{j=0}^{\infty}\left|\widetilde{\mathbf{B}}_{j,t}\right| \le \max_{t\ge 1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}|\mathbf{B}_{k,t}| = \max_{t\ge 1}\sum_{j=1}^{\infty} j\,|\mathbf{B}_{j,t}| < \infty.
$$

(4). By Assumption 2.2.1,

$$
\begin{aligned}
\sum_{t=1}^{T-1}|\widetilde{\mathbb{B}}_{t+1}(1) - \widetilde{\mathbb{B}}_t(1)| \;\; & \le \;\; \sum_{t=1}^{T-1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}|\mathbf{B}_{k,t+1} - \mathbf{B}_{k,t}| \\
& = \;\; \sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\,|\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}| < \infty.
\end{aligned}
$$

(5). By Assumption 2.2.1,

$$
\begin{aligned}
\max_{t\ge 1}\sum_{j=0}^{\infty}\left|\widetilde{\mathbf{B}}_{j,t}^r\right| & \le \max_{t\ge 1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}|\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t}| = \max_{t\ge 1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}|\mathbf{B}_{k+r,t}|\cdot|\mathbf{B}_{k,t}| \\
& = \max_{t\ge 1}\sum_{j=1}^{\infty} j\,|\mathbf{B}_{j+r,t}|\cdot|\mathbf{B}_{j,t}| \le M\max_{t\ge 1}\sum_{j=1}^{\infty} j\,|\mathbf{B}_{j,t}| < \infty.
\end{aligned}
$$

(6). Write

$$
\begin{aligned}
& \max_{t\ge 1}\sum_{r=1}^{\infty}\left|\widetilde{\mathbb{B}}_t^r(1)\right| \le \max_{t\ge 1}\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}|\mathbf{B}_{k+r,t}|\cdot|\mathbf{B}_{k,t}| \\
& = \max_{t\ge 1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}|\mathbf{B}_{k,t}|\cdot\left(\sum_{r=1}^{\infty}|\mathbf{B}_{k+r,t}|\right) \le \max_{t\ge 1}\left(\sum_{r=1}^{\infty}|\mathbf{B}_{r,t}|\right)\cdot\left(\sum_{j=1}^{\infty} j\,|\mathbf{B}_{j,t}|\right) < \infty.
\end{aligned}
$$

(7). Write

$$\begin{aligned}
&\sum_{t=1}^{T-1}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\left|\widetilde{\mathbb{B}}_{t+1}^{r}(1)-\widetilde{\mathbb{B}}_{t}^{r}(1)\right| \le \sum_{t=1}^{T-1}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\left|\sum_{k=j+1}^{\infty}\mathbf{B}_{k+r,t+1}\otimes\mathbf{B}_{k,t+1}-\mathbf{B}_{k+r,t}\otimes\mathbf{B}_{k,t}\right| \\
\le\ & \sum_{t=1}^{T-1}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\left(|\mathbf{B}_{k+r,t+1}-\mathbf{B}_{k+r,t}|\cdot|\mathbf{B}_{k,t+1}|+|\mathbf{B}_{k+r,t}|\cdot|\mathbf{B}_{k,t+1}-\mathbf{B}_{k,t}|\right) \\
=\ & \sum_{t=1}^{T-1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\left(|\mathbf{B}_{k,t+1}|\cdot\sum_{r=0}^{\infty}|\mathbf{B}_{k+r,t+1}-\mathbf{B}_{k+r,t}|+|\mathbf{B}_{k,t+1}-\mathbf{B}_{k,t}|\cdot\sum_{r=0}^{\infty}|\mathbf{B}_{k+r,t}|\right) \\
\le\ & \left(\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}|\mathbf{B}_{r,t+1}-\mathbf{B}_{r,t}|\right)\cdot\left(\max_{t\ge1}\sum_{k=1}^{\infty}k\,|\mathbf{B}_{k,t}|\right) \\
&+\left(\sum_{t=1}^{T-1}\sum_{k=1}^{\infty}k\,|\mathbf{B}_{k,t+1}-\mathbf{B}_{k,t}|\right)\cdot\left(\max_{t\ge1}\sum_{r=1}^{\infty}|\mathbf{B}_{r,t}|\right)<\infty.
\end{aligned}$$

The proof is now completed. □

Proof of Lemma A.1.4.

In the following proof, I cover the interval $[a,b]$ by a finite number of subintervals $\{S_l\}$, which are centered at s_l with the length denoted by δ_T. Denoting the number of these intervals by N_T, then $N_T=O(\delta_T^{-1})$. In addition, let $\delta_T=O(T^{-1}\gamma_T)$ with $\gamma_T=\sqrt{d_T\log T}$.

Write

$$\begin{aligned}
&\sup_{\tau\in[a,b]}\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}(\tau)\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t\right| \le \max_{1\le l\le N_T}\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}(s_l)\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t\right| \\
&+\max_{1\le l\le N_T}\sup_{\tau\in S_l}\left|\sum_{t=1}^{T}\left(\mathbf{W}_{T,t}(\tau)-\mathbf{W}_{T,t}(s_l)\right)\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t\right| := J_{T,1}+J_{T,2}.
\end{aligned}$$

For $J_{T,2}$, since $\mathbf{W}_{T,t}(\cdot)$ is Lipschitz continuous and $\max_{t\ge1}|\mathbb{B}_t(1)|<\infty$ by Assumption 2.2.1, I have

$$\begin{aligned}
E|J_{T,2}| &\le \sum_{t=1}^{T}\max_{1\le l\le N_T}\sup_{\tau\in S_l}|\mathbf{W}_{T,t}(\tau)-\mathbf{W}_{T,t}(s_l)|\,E\,|\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t| \\
&\le MT\delta_T\max_{t\ge1}E|\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t| = O(\gamma_T).
\end{aligned}$$

For $J_{T,1}$, I apply the truncation method. Define $\boldsymbol{\varepsilon}_t'=\boldsymbol{\varepsilon}_t I(|\boldsymbol{\varepsilon}_t|\le T^{\frac{1}{\delta}})$ and $\boldsymbol{\varepsilon}_t''=\boldsymbol{\varepsilon}_t-\boldsymbol{\varepsilon}_t'$, where δ is defined in Assumption 2.2.2, and $I(\cdot)$ is the indicator function. Write

$$\begin{aligned}
J_{T,1} &= \max_{1\le l\le N_T}\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}(s_l)\mathbb{B}_t(1)\left(\boldsymbol{\varepsilon}_t'+\boldsymbol{\varepsilon}_t''-E\left(\boldsymbol{\varepsilon}_t'+\boldsymbol{\varepsilon}_t''|\mathscr{F}_{t-1}\right)\right)\right| \\
&\le \max_{1\le l\le N_T}\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}(s_l)\mathbb{B}_t(1)\left(\boldsymbol{\varepsilon}_t'-E(\boldsymbol{\varepsilon}_t'|\mathscr{F}_{t-1})\right)\right|+\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}(s_l)\mathbb{B}_t(1)\boldsymbol{\varepsilon}_t''\right| \\
&+\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}\mathbf{W}_{T,t}(s_l)\mathbb{B}_t(1)E(\boldsymbol{\varepsilon}_t''|\mathscr{F}_{t-1})\right| := J_{T,11}+J_{T,12}+J_{T,13}.
\end{aligned}$$

Start from $J_{T,12}$. By Hölder's inequality and Markov's inequality,

$$
\begin{aligned}
E|J_{T,12}| &\leq O(1)d_T\sum_{t=1}^T E\left|\varepsilon_t''\right| = O(1)d_T\sum_{t=1}^T E|\varepsilon_t I(|\varepsilon_t|\geq T^{\frac{1}{\delta}})| \\
&\leq O(1)d_T\sum_{t=1}^T\left\{E|\varepsilon_t|^\delta\right\}^{\frac{1}{\delta}}\left\{E|I(|\varepsilon_t|\geq T^{\frac{1}{\delta}})|\right\}^{\frac{\delta-1}{\delta}} \\
&= O(1)d_T\sum_{t=1}^T\left\{E|\varepsilon_t|^\delta\right\}^{\frac{1}{\delta}}\left\{\Pr\left(|\varepsilon_t|\geq T^{\frac{1}{\delta}}\right)\right\}^{\frac{\delta-1}{\delta}} \\
&\leq O(1)d_T\sum_{t=1}^T\left\{E|\varepsilon_t|^\delta\right\}^{\frac{1}{\delta}}\left\{\frac{E|\varepsilon_t|^\delta}{T}\right\}^{\frac{\delta-1}{\delta}} = O(T^{\frac{1}{\delta}}d_T) = o\left(\sqrt{d_T\log T}\right),
\end{aligned}
$$

where the second inequality follows from Hölder's inequality, and the third inequality follows from Markov's inequality. Similarly, $J_{T,13} = O_P(T^{\frac{1}{\delta}}d_T) = o_P\left(\sqrt{d_T\log T}\right)$.

I now turn to $J_{T,11}$. For notational simplicity, let $\mathbf{Y}_t = \mathbf{W}_{T,t}(s_l)\mathbb{B}_t(1)\left(\varepsilon_t' - E(\varepsilon_t'|\mathscr{F}_{t-1})\right)$ for $1\leq t\leq T$ and $A_T = 2T^{\frac{1}{\delta}}d_T\max_{t\geq 1}|\mathbb{B}_t(1)|$. Simple algebra shows that $|\mathbf{Y}_t|\leq A_T$ uniformly in t and s_l. By Assumption 2.2.2 and the first condition in the body of this lemma,

$$
\max_{1\leq l\leq N_T}\sum_{t=1}^T E\left(|\mathbf{Y}_t|^2|\mathscr{F}_{t-1}\right) \leq Md_T\max_{1\leq l\leq N_T}\sum_{t=1}^T|\mathbf{W}_{T,t}(s_l)|\,E\left(|\varepsilon_t|^2|\mathscr{F}_{t-1}\right) = O_{a.s.}(d_T).
$$

By Lemma A.1.2 and $T^{\frac{2}{\delta}}d_T\log T\to 0$, choose some $\beta>0$ (such as $\beta=4$), and write

$$
\begin{aligned}
&\Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T\right) \\
= &\Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T,\ \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|\leq Md_T\right) \\
&+\Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T,\ \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|> Md_T\right) \\
\leq &\Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T,\ \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|\leq Md_T\right) \\
&+\Pr\left(\max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|> Md_T\right) \\
\leq &N_T\exp\left(-\frac{\beta M\gamma_T^2}{2(Md_T+\gamma_T 2A_T)}\right)\leq N_T\exp\left(-\frac{\beta}{2}\log T\right) = O(\delta_T^{-1})T^{-\frac{\beta}{2}}\to 0.
\end{aligned}
$$

Based on the above development, the proof is now completed. □

Proof of Lemma A.1.5.

(1). Similar to the proof of Lemma A.1.4, I use a finite number of subintervals $\{S_l\}$ to cover the interval $[a,b]$, which are centered at s_l with the length δ_T. Denote the number of these intervals by N_T then $N_T = O(\delta_T^{-1})$. In addition, let $\delta_T = O(T^{-1}\gamma_T)$ with $\gamma_T = \sqrt{d_T \log T}$.

$$\begin{aligned}
&\sup_{\tau\in[a,b]} \left| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(\tau)) \mathbb{B}_t^0(1) \left(\text{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\right) - \text{vec}(\mathbf{I}_d)\right) \right| \\
\leq & \max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(s_l)) \mathbb{B}_t^0(1) \left(\text{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\right) - \text{vec}(\mathbf{I}_d)\right) \right| \\
& + \max_{1\leq l\leq N_T} \sup_{\tau\in S_l} \left| \sum_{t=1}^{T} \left(\mathbf{I}_d \otimes (\mathbf{W}_{T,t}(\tau) - \mathbf{W}_{T,t}(s_l))\right) \mathbb{B}_t^0(1) \left(\text{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\right) - \text{vec}(\mathbf{I}_d)\right) \right| \\
:= & J_{T,1} + J_{T,2}.
\end{aligned}$$

Start from $J_{T,2}$. Similar to the proof of Lemma A.1.4, since

$$\left|\mathbb{B}_t^0(1)\right| \leq \sum_{j=0}^{\infty} |\mathbf{B}_{j,t}|^2 \leq \left(\sum_{j=0}^{\infty} |\mathbf{B}_{j,t}|\right)^2 < \infty$$

by Assumption 2.2.1, I have

$$E|J_{T,2}| \leq MT\delta_T \max_{t\geq 1} E\left|\mathbb{B}_t^0(1)\left(\text{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\right) - \text{vec}(\mathbf{I}_d)\right)\right| = O(\gamma_T).$$

I then apply the truncation method. Define $\mathbf{u}_t = \mathbb{B}_t^0(1)\left(\text{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\right) - \text{vec}\left(\mathbf{I}_d\right)\right)$, $\mathbf{u}_t' = \mathbf{u}_t I(|\mathbf{u}_t| \leq T^{\frac{2}{\delta}})$ and $\mathbf{u}_t'' = \mathbf{u}_t - \mathbf{u}_t'$. For $J_{T,1}$, write

$$\begin{aligned}
J_{T,1} = & \max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(s_l)) \left(\mathbf{u}_t' + \mathbf{u}_t'' - E(\mathbf{u}_t' + \mathbf{u}_t''|\mathscr{F}_{t-1})\right) \right| \\
\leq & \max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(s_l)) \left(\mathbf{u}_t' - E(\mathbf{u}_t'|\mathscr{F}_{t-1})\right) \right| + \max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(s_l)) \mathbf{u}_t'' \right| \\
& + \max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(s_l)) E(\mathbf{u}_t''|\mathscr{F}_{t-1}) \right| := J_{T,11} + J_{T,12} + J_{T,13}.
\end{aligned}$$

As in the proof of Lemma A.1.4, I can show that $J_{T,12} = o_P(\sqrt{d_T \log T})$ and $J_{T,13} = o_P(\sqrt{d_T \log T})$ respectively. I focus on $J_{T,11}$ below.

For any $1 \leq l \leq N_T$, let $\mathbf{Y}_t = (\mathbf{I}_d \otimes \mathbf{W}_{T,t}(s_l))(\mathbf{u}_t' - E(\mathbf{u}_t'|\mathscr{F}_{t-1}))$. I then have $E\left(\mathbf{Y}_t|\mathscr{F}_{t-1}\right) = 0$ and $|\mathbf{Y}_t| \leq 2T^{2/\delta} d_T \max_t \left|\mathbb{B}_t^0(1)\right|$. Since $\max_{t\geq 1} E\left(|\boldsymbol{\varepsilon}_t|^4 |\mathscr{F}_{t-1}\right) < \infty$ a.s., I can write

$$\max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} E(\mathbf{Y}_t \mathbf{Y}_t^\top|\mathscr{F}_{t-1}) \right| \leq M d_T \max_{t\geq 1} E\left(|\mathbf{u}_t|^2 |\mathscr{F}_{t-1}\right) \max_{1\leq l\leq N_T} \sum_{t=1}^{T} |\mathbf{W}_{T,t}(s_l)| = O_{a.s.}\left(d_T\right).$$

Similar to Lemma A.1.2, choose $\beta > 0$ (such as $\beta = 4$). In view of the fact that $T^{\frac{4}{\delta}} d_T \log T \to 0$, I write

$$\begin{aligned}\Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T\right) &= \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| \le M d_T\right)\\ &\quad + \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| > M d_T\right)\\ &\le \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| \le M d_T\right)\\ &\quad + \Pr\left(\max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| > M d_T\right)\\ &\le N_T \exp\left(-\frac{\beta M\gamma_T^2}{2(Md_T + M\gamma_T T^{\frac{2}{\delta}} d_T)}\right)\\ &\le N_T\exp\left(-\frac{\beta}{2}\log T\right) = N_T T^{-\frac{\beta}{2}} = o(1).\end{aligned}$$

The first result then follows.

(2). Let $\{S_l\}$ be a finite number of subintervals covering the interval $[a,b]$, which are centered at s_l with the length δ_T. Denote the number of these intervals by N_T then $N_T = O(\delta_T^{-1})$. In addition, let $\delta_T = O(T^{-1}\gamma_T)$ with $\gamma_T = \sqrt{d_T\log T}$. Then

$$\begin{aligned}\sup_{\tau\in[a,b]}\left|\sum_{t=1}^T(\mathbf{I}_d\otimes\mathbf{W}_{T,t}(\tau))\,\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right| &\le \max_{1\le l\le N_T}\left|\sum_{t=1}^T(\mathbf{I}_d\otimes\mathbf{W}_{T,t}(s_l))\,\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|\\ + \max_{1\le l\le N_T}\sup_{\tau\in S_l}\left|\sum_{t=1}^T(\mathbf{I}_d\otimes(\mathbf{W}_{T,t}(\tau)-\mathbf{W}_{T,t}(s_l)))\,\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right| &:= J_{T,3} + J_{T,4}.\end{aligned}$$

Consider $J_{T,4}$. By the fact that $|\mathrm{tr}(\mathbf{A})| \le d\,|\mathbf{A}|$ for any $d \times d$ matrix $\mathbf{A}$ and Assumption 2.2.1,

$$\begin{aligned}
E\left|\boldsymbol{\zeta}_t \boldsymbol{\varepsilon}_t\right| &= E\left|\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\mathbf{B}_{s+r,t}\otimes\mathbf{B}_{s,t}\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-r}^{\top}\right)\right| \\
&\le \left(E\left|\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\mathbf{B}_{s+r,t}\otimes\mathbf{B}_{s,t}\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-r}^{\top}\right)\right|^2\right)^{1/2} \\
&\le \left\{\mathrm{tr}\left[\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\mathbf{B}_{s+r,t}\otimes\mathbf{B}_{s,t}\right)\cdot(\mathbf{I}_d\otimes\mathbf{I}_d)\cdot\left(\sum_{s=0}^{\infty}\mathbf{B}_{s+r,t}\otimes\mathbf{B}_{s,t}\right)^{\top}\right]\right\}^{1/2} \\
&\le M\left(\sum_{r=1}^{\infty}\left|\sum_{s=0}^{\infty}\mathbf{B}_{s+r,t}\otimes\mathbf{B}_{s,t}\right|^2\right)^{1/2} \\
&\le M\left(\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\left|\mathbf{B}_{s+r,t}\right|^2\right)\cdot\left(\sum_{s=0}^{\infty}\left|\mathbf{B}_{s,t}\right|^2\right)\right)^{1/2} \\
&= M\left(\left(\sum_{r=1}^{\infty}r\left|\mathbf{B}_{r,t}\right|^2\right)\cdot\left(\sum_{s=0}^{\infty}\left|\mathbf{B}_{s,t}\right|^2\right)\right)^{1/2} \\
&\le M\left(\left(\sum_{r=1}^{\infty}r\left|\mathbf{B}_{r,t}\right|\right)^2\cdot\left(\sum_{s=0}^{\infty}\left|\mathbf{B}_{s,t}\right|^2\right)\right)^{1/2} < \infty.
\end{aligned}$$

Similarly, I have $E|J_{T,4}| \le MT\delta_T \max_{t\ge 1} E\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right| = O(\gamma_T)$.

Before investigating $J_{T,3}$, I first show that

$$\max_{1\le l\le N_T}\sum_{t=1}^{T}\left|\mathbf{W}_{T,t}(s_l)\right| E\left(\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2|\mathscr{F}_{t-1}\right) = O_P(1). \tag{A.3.1}$$

Note that

$$\begin{aligned}
&\max_{1\le l\le N_T}\sum_{t=1}^{T}\left|\mathbf{W}_{T,t}(s_l)\right| E\left(\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2|\mathscr{F}_{t-1}\right) \\
\le\ &\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}\left|\mathbf{W}_{T,t}(s_l)\right|\left(E\left(\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2|\mathscr{F}_{t-1}\right)-E\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2\right)\right| \\
&+\max_{1\le l\le N_T}\sum_{t=1}^{T}\left|\mathbf{W}_{T,t}(s_l)\right| E\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2
\end{aligned}$$

and $\max_{1\le l\le N_T}\sum_{t=1}^{T}\left|\mathbf{W}_{T,t}(s_l)\right| E\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2 = O(1)$. Thus, to prove (A.3.1), it is sufficient to show

$$\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}\left|\mathbf{W}_{T,t}(s_l)\right|\left(E\left(\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2|\mathscr{F}_{t-1}\right)-E\left|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t\right|^2\right)\right| = o_P(1).$$

In order to do so, I write

$$\begin{aligned}
&\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\left(E\left(|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t|^2\,|\mathscr{F}_{t-1}\right)-E\,|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t|^2\right)\right|\\
=&\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\,\mathrm{tr}\left(\sum_{r,r^*=1}^{\infty}\mathbb{B}_t^r(1)(\boldsymbol{\varepsilon}_{t-r}\boldsymbol{\varepsilon}_{t-r^*}^\top\otimes\mathbf{I}_d)\mathbb{B}_t^{r^*,\top}(1)-\sum_{r=1}^{\infty}\mathbb{B}_t^r(1)\mathbb{B}_t^{r,\top}(1)\right)\right|\\
\le\,& d^2\cdot\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\sum_{r=1}^{\infty}(\mathbb{B}_t^r(1)\otimes\mathbb{B}_t^r(1))\left(\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t-r}\boldsymbol{\varepsilon}_{t-r}^\top\otimes\mathbf{I}_d\right)-\mathrm{vec}(\mathbf{I}_{d^2})\right)\right|\\
&+2d^2\cdot\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\left(\mathbb{B}_t^{r+j}(1)\otimes\mathbb{B}_t^r(1)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t-r}\boldsymbol{\varepsilon}_{t-r-j}^\top\otimes\mathbf{I}_d\right)\right|\\
:=&\, J_{T,5}+J_{T,6}.
\end{aligned}$$

Let $\mathbb{F}_{r,t}(L)=\sum_{j=1}^{\infty}\mathbb{B}_t^{r+j}(1)\otimes\mathbb{B}_t^j(1)L^j$. Similar to the second result of Lemma A.1.3, I have

$$\mathbb{F}_{r,t}(L)=\mathbb{F}_{r,t}(1)-(1-L)\widetilde{\mathbb{F}}_{r,t}(L) \tag{A.3.2}$$

where $\widetilde{\mathbb{F}}_{r,t}(L)=\sum_{j=1}^{\infty}\widetilde{\mathbb{F}}_{rj,t}L^j$ and $\widetilde{\mathbb{F}}_{rj,t}=\sum_{k=j+1}^{\infty}\mathbb{B}_t^{r+k}(1)\otimes\mathbb{B}_t^k(1)$. For notational simplicity, denote

$$\begin{aligned}
\widetilde{X}_{at}&=\sum_{j=1}^{\infty}\left(\mathbb{B}_t^j(1)\otimes\mathbb{B}_t^j(1)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-j}^\top\otimes\mathbf{I}_d\right),\\
\widetilde{X}_{bt}&=\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\left(\mathbb{B}_t^{r+j}(1)\otimes\mathbb{B}_t^j(1)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-r-j}^\top\otimes\mathbf{I}_d\right).
\end{aligned}$$

Applying (A.3.2) to $\widetilde{X}_{at}$ and $\widetilde{X}_{bt}$ yields that

$$\begin{aligned}
\widetilde{X}_{at}&=\mathbb{F}_{0,t}(1)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\mathbf{I}_d\right)-(1-L)\widetilde{\mathbb{F}}_{0,t}(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\mathbf{I}_d\right),\\
\widetilde{X}_{bt}&=\sum_{r=1}^{\infty}\mathbb{F}_{r,t}(1)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-r}^\top\otimes\mathbf{I}_d\right)-(1-L)\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,t}(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-r}^\top\otimes\mathbf{I}_d\right).
\end{aligned}$$

For $J_{T,5}$, summing up $\widetilde{X}_{at}$ over t yields

$$\begin{aligned}
&\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\sum_{r=1}^{\infty}(\mathbb{B}_t^r(1)\otimes\mathbb{B}_t^r(1))\left(\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t-r}\boldsymbol{\varepsilon}_{t-r}^\top\otimes\mathbf{I}_d\right)-\mathrm{vec}(\mathbf{I}_{d^2})\right)\right|\\
\le\;&\max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\,\mathbb{F}_{0,t}(1)\left(\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\mathbf{I}_d\right)-\mathrm{vec}\left(\mathbf{I}_{d^2}\right)\right)\right|\\
&+\max_{1\le l\le N_T}\left||\mathbf{W}_{T,1}(s_l)|\,\widetilde{\mathbb{F}}_{0,1}(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_0\boldsymbol{\varepsilon}_0^\top\otimes\mathbf{I}_d\right)\right|\\
&+\sup_{0\le\tau\le1}\left||\mathbf{W}_{T,T}(s_l)|\,\widetilde{\mathbb{F}}_{0,T}(L)\mathrm{vec}\left(\boldsymbol{\varepsilon}_T\boldsymbol{\varepsilon}_T^\top\otimes\mathbf{I}_d\right)\right|\\
&+\max_{1\le l\le N_T}\left|\sum_{t=1}^{T-1}\left(|\mathbf{W}_{T,t+1}(s_l)|\,\widetilde{\mathbb{F}}_{0,t+1}(L)-|\mathbf{W}_{T,t}(s_l)|\,\widetilde{\mathbb{F}}_{0,t}(L)\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\mathbf{I}_d\right)\right|\\
:=\;&J_{T,51}+J_{T,52}+J_{T,53}+J_{T,54}.
\end{aligned}$$

Similar to the proof of Lemma A.1.5.1, I can show that $J_{T,51} = O_P\left(\sqrt{d_T \log T}\right)$, since

$$\max_{t\geq 1} |\mathbb{F}_{0,t}(1)| \leq \max_{t\geq 1} \sum_{j=1}^{\infty} \left| \sum_{k=0}^{\infty} \mathbf{B}_{k+j,t} \otimes \mathbf{B}_{k,t} \right|^2 \leq \max_{t\geq 1} \sum_{j=1}^{\infty} \left(\sum_{k=0}^{\infty} |\mathbf{B}_{k+j,t}|^2 \right) \left(\sum_{k=0}^{\infty} |\mathbf{B}_{k,t}|^2 \right)$$
$$\leq \max_{t\geq 1} \left(\sum_{k=0}^{\infty} |\mathbf{B}_{k,t}|^2 \right) \left(\sum_{j=1}^{\infty} j\, |\mathbf{B}_{j,t}|^2 \right) < \infty.$$

Also, I can show that $J_{T,52} = O_P(d_T)$ and $J_{T,53} = O_P(d_T)$, since

$$\begin{aligned}
\max_{t\geq 1} \left| \widetilde{\mathbb{F}}_{0,t}(1) \right| &\leq \max_{t\geq 1} \sum_{j=1}^{\infty} \sum_{k=j+1}^{\infty} \left| \mathbb{B}_t^k(1) \right|^2 \leq \max_{t\geq 1} \sum_{r=1}^{\infty} \sum_{k=r+1}^{\infty} \left(\sum_{j=0}^{\infty} |\mathbf{B}_{j+k,t}|^2 \right) \left(\sum_{j=0}^{\infty} |\mathbf{B}_{j,t}|^2 \right) \\
&\leq \max_{t\geq 1} \left(\sum_{j=0}^{\infty} |\mathbf{B}_{j,t}|^2 \right) \left(\sum_{r=1}^{\infty} \sum_{k=r+1}^{\infty} (k-r)\, |\mathbf{B}_{k,t}|^2 \right) \\
&\leq \max_{t\geq 1} \left(\sum_{j=0}^{\infty} |\mathbf{B}_{j,t}|^2 \right) \left(\sum_{r=1}^{\infty} \frac{r(r+1)}{2}\, |\mathbf{B}_{r+1,t}|^2 \right) \\
&\leq \max_{t\geq 1} \left(\sum_{j=0}^{\infty} |\mathbf{B}_{j,t}|^2 \right) \left(\sum_{j=1}^{\infty} j^2\, |\mathbf{B}_{j,t}|^2 \right) < \infty.
\end{aligned}$$

I can easily show $J_{T,54} = o_P(1)$, since

$$\sup_{\tau\in[a,b]} \left(\sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1}(\tau)| - |\mathbf{W}_{T,t}(\tau)| \right) \leq \sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} |\mathbf{W}_{T,t+1}(\tau) - \mathbf{W}_{T,t}(\tau)| = o(1)$$

and

$$\begin{aligned}
&\sum_{t=1}^{T-1} \left| \widetilde{\mathbb{F}}_{0,t+1}(1) - \widetilde{\mathbb{F}}_{0,t}(1) \right| \leq \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \sum_{k=r+1}^{\infty} \left| \mathbb{B}_{t+1}^k(1) \otimes \mathbb{B}_{t+1}^k(1) - \mathbb{B}_t^k(1) \otimes \mathbb{B}_t^k(1) \right| \\
\leq\ & \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \sum_{k=r+1}^{\infty} \left| \mathbb{B}_{t+1}^k(1) - \mathbb{B}_t^k(1) \right| \cdot \left(\left| \mathbb{B}_{t+1}^k(1) \right| + \left| \mathbb{B}_t^k(1) \right| \right) \\
\leq\ & M \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \sum_{k=r+1}^{\infty} \sum_{j=0}^{\infty} |\mathbf{B}_{j+k,t+1} \otimes \mathbf{B}_{j,t+1} - \mathbf{B}_{j+k,t} \otimes \mathbf{B}_{j,t}| \\
\leq\ & M \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \sum_{k=r+1}^{\infty} \sum_{j=0}^{\infty} \left(|\mathbf{B}_{j+k,t+1} - \mathbf{B}_{j+k,t}| \cdot |\mathbf{B}_{j,t+1}| + |\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}| \cdot |\mathbf{B}_{j+k,t}| \right) \\
\leq\ & M \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \sum_{k=r+1}^{\infty} \left(|\mathbf{B}_{k,t+1} - \mathbf{B}_{k,t}| \cdot \sum_{j=0}^{\infty} |\mathbf{B}_{j,t+1}| + |\mathbf{B}_{k,t}| \cdot \sum_{j=0}^{\infty} |\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}| \right) \\
\leq\ & M \left(\sum_{t=1}^{T-1} \sum_{k=1}^{\infty} k\, |\mathbf{B}_{k,t+1} - \mathbf{B}_{k,t}| \right) \cdot \left(\max_{t\geq 1} \sum_{j=0}^{\infty} |\mathbf{B}_{j,t+1}| \right) \\
& + M \left(\max_{t\geq 1} \sum_{k=1}^{\infty} k\, |\mathbf{B}_{k,t}| \right) \cdot \left(\sum_{t=1}^{T-1} \sum_{j=0}^{\infty} |\mathbf{B}_{j,t+1} - \mathbf{B}_{j,t}| \right) = O(1).
\end{aligned}$$

Based on the above development, I conclude that $J_{T,5} = o_P(1)$. Next, I focus on $J_{T,6}$, and write

$$\begin{aligned}
& \max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\mathbb{B}_t^{r+j}(1)\otimes\mathbb{B}_t^{r}(1)\text{vec}\left(\boldsymbol{\varepsilon}_{t-r}\boldsymbol{\varepsilon}_{t-r-j}^\top\otimes\mathbf{I}_d\right)\right| \\
\le\ & \max_{1\le l\le N_T}\left|\sum_{t=1}^{T}|\mathbf{W}_{T,t}(s_l)|\sum_{r=1}^{\infty}\mathbb{F}_{r,t}(1)\text{vec}\left(\boldsymbol{\varepsilon}_{t}\boldsymbol{\varepsilon}_{t-r}^\top\otimes\mathbf{I}_d\right)\right| \\
& +\max_{1\le l\le N_T}\left||\mathbf{W}_{T,1}(s_l)|\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,1}(L)\text{vec}\left(\boldsymbol{\varepsilon}_{0}\boldsymbol{\varepsilon}_{-r}^\top\otimes\mathbf{I}_d\right)\right| \\
& +\max_{1\le l\le N_T}\left||\mathbf{W}_{T,T}(s_l)|\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,T}(L)\text{vec}\left(\boldsymbol{\varepsilon}_{t}\boldsymbol{\varepsilon}_{T-r}^\top\otimes\mathbf{I}_d\right)\right| \\
& +\max_{1\le l\le N_T}\left|\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left(|\mathbf{W}_{T,t+1}(s_l)|\,\widetilde{\mathbb{F}}_{r,t+1}(L)-|\mathbf{W}_{T,t}(s_l)|\,\widetilde{\mathbb{F}}_{r,t}(L)\right)\text{vec}\left(\boldsymbol{\varepsilon}_{t}\boldsymbol{\varepsilon}_{t-r}^\top\otimes\mathbf{I}_d\right)\right| \\
:=\ & J_{T,61}+J_{T,62}+J_{T,63}+J_{T,64}.
\end{aligned}$$

I can show that $J_{T,62}$ and $J_{T,63}$ are $O_P(d_T)$, since

$$\begin{aligned}
& \max_{t\ge1}\left|\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,t}(1)\right| \le \max_{t\ge1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\left|\widetilde{\mathbb{F}}_{rj,t}\right| \le \max_{t\ge1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left|\mathbb{B}_t^{r+k}(1)\right|\left|\mathbb{B}_t^{k}(1)\right| \\
\le\ & \max_{t\ge1}\left(\sum_{r=1}^{\infty}|\mathbb{B}_t^{r}(1)|\right)\left(\sum_{j=1}^{\infty}j\left|\mathbb{B}_t^{j}(1)\right|\right) \le M\max_{t\ge1}\sum_{j=1}^{\infty}j\sum_{k=0}^{\infty}|\mathbf{B}_{k+j,t}|\,|\mathbf{B}_{k,t}| \\
\le\ & M\max_{t\ge1}\left(\sum_{j=1}^{\infty}j\,|\mathbf{B}_{j,t}|\right)\left(\sum_{k=0}^{\infty}|\mathbf{B}_{k,t}|\right)<\infty.
\end{aligned}$$

Similar to $J_{T,54}$, I have $J_{T,64} = o_P\left(1\right)$, since

$$\begin{aligned}
& \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left|\widetilde{\mathbb{F}}_{r,t+1}(1)-\widetilde{\mathbb{F}}_{r,t}(1)\right| \\
\le\ & \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left|\mathbb{B}_{t+1}^{r+k}(1)\otimes\mathbb{B}_{t+1}^{k}(1)-\mathbb{B}_{t}^{r+k}(1)\otimes\mathbb{B}_{t}^{k}(1)\right| \\
\le\ & \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left(\left|\mathbb{B}_{t+1}^{r+k}(1)-\mathbb{B}_{t}^{r+k}(1)\right|\left|\mathbb{B}_{t+1}^{k}(1)\right|+\left|\mathbb{B}_{t+1}^{k}(1)-\mathbb{B}_{t}^{k}(1)\right|\left|\mathbb{B}_{t}^{r+k}(1)\right|\right) \\
\le\ & \left(\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left|\mathbb{B}_{t+1}^{r}(1)-\mathbb{B}_{t}^{r}(1)\right|\right)\left(\max_{t\ge1}\sum_{j=1}^{\infty}j\left|\mathbb{B}_t^{j}\right|\right) \\
& +\left(\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\left|\mathbb{B}_{t+1}^{j}(1)-\mathbb{B}_{t}^{j}(1)\right|\right)\left(\max_{t\ge1}\sum_{r=1}^{\infty}|\mathbb{B}_t^{r}|\right)=O(1).
\end{aligned}$$

Now consider term $J_{T,61}$. Define $\mathbf{u}_t = \sum_{r=1}^{\infty}\mathbb{F}_{r,t}(1)\text{vec}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-r}^\top\otimes\mathbf{I}_d\right)$, $\mathbf{u}_t' = \mathbf{u}_t I(|\mathbf{u}_t|\le T^{\frac{2}{\delta}})$ and

$\mathbf{u}_t'' = \mathbf{u}_t - \mathbf{u}_t'$. Then I have

$$\begin{aligned} J_{T,61} = & \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T |\mathbf{W}_{T,t}(s_l)|\left(\mathbf{u}_t' + \mathbf{u}_t'' - E(\mathbf{u}_t' + \mathbf{u}_t''|\mathscr{F}_{t-1})\right)\right| \\ \leq & \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T |\mathbf{W}_{T,t}(s_l)|\left(\mathbf{u}_t' - E(\mathbf{u}_t'|\mathscr{F}_{t-1})\right)\right| + \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T |\mathbf{W}_{T,t}(s_l)|\,\mathbf{u}_t''\right| \\ & + \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T |\mathbf{W}_{T,t}(s_l)|\,E(\mathbf{u}_t''|\mathscr{F}_{t-1})\right| := J_{T,611} + J_{T,612} + J_{T,613}. \end{aligned}$$

Using the same argument as that used in the proof of $J_{T,12}$ in Lemma A.1.4, I can show that $J_{T,612}$ and $J_{T,613}$ are $O_P\left(T^{\frac{2}{\delta}}d_T\right)$. Next, consider $J_{T,611}$. For any $1 \leq l \leq N_T$, let $\mathbf{Y}_t = |\mathbf{W}_{T,t}(s_l)|\,(\mathbf{u}_t' - E(\mathbf{u}_t'|\mathscr{F}_{t-1}))$. I then have $E\left(\mathbf{Y}_t|\mathscr{F}_{t-1}\right) = 0$ and $|\mathbf{Y}_t| \leq 2T^{2/\delta}d_T$. In addition, I have

$$\begin{aligned} & \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| \leq 4\max_{1\leq l\leq N_T}\sum_{t=1}^T |\mathbf{W}_{T,t}(s_l)|^2\,E\left(|\mathbf{u}_t|^2\,|\mathscr{F}_{t-1}\right) \\ \leq & \ M\cdot d_T\max_{1\leq l\leq N_T}\sum_{t=1}^T |\mathbf{W}_{T,t}(s_l)|\sum_{r=1}^\infty |\mathbb{F}_{r,t}(1)|^2\,|\boldsymbol{\varepsilon}_{t-r}|^2 \leq M\cdot d_T\max_{t\geq 1}\sum_{r=1}^\infty |\mathbb{F}_{r,t}(1)|^2\,|\boldsymbol{\varepsilon}_{t-r}|^2 \\ \leq & \ M\cdot d_T\sum_{r=1}^\infty \max_{t\geq 1}|\mathbb{F}_{r,t}(1)|^2\left(\sum_{t=1}^T|\boldsymbol{\varepsilon}_{t-r}|^\delta\right)^{\frac{2}{\delta}} = O_P\left(d_T T^{\frac{2}{\delta}}\right). \end{aligned}$$

Therefore, I have $\max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| = O_P(d_T T^{\frac{2}{\delta}})$. By Lemma A.1.2, and choosing $\beta = 4$, I have

$$\begin{aligned} & \Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T}\right) \\ = & \ \Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T},\ \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| \leq M d_T T^{\frac{2}{\delta}}\right) \\ & +\Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T},\ \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| > M d_T T^{\frac{2}{\delta}}\right) \\ \leq & \ \Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T},\ \max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| \leq M d_T T^{\frac{2}{\delta}}\right) \\ & +\Pr\left(\max_{1\leq l\leq N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| > M d_T T^{\frac{2}{\delta}}\right) \\ \leq & \ N_T\exp\left(-\frac{\beta M d_T T^{\frac{2}{\delta}}\log T}{2(M d_T T^{\frac{2}{\delta}} + M\sqrt{d_T T^{\frac{2}{\delta}}\log T}T^{\frac{2}{\delta}}d_T)}\right) + o(1) \\ \leq & \ N_T\exp\left(-\frac{\beta}{2}\log T\right) = N_T T^{-\frac{\beta}{2}} = o(1) \end{aligned}$$

given $d_T T^{\frac{4}{\delta}}\log T \to 0$. Hence, I have $J_{T,611} = O_P(\{d_T T^{\frac{2}{\delta}}\log T\}^{1/2})$. Combining the above results, I have proved that $\sup_{\tau\in[a,b]}\left|\sum_{t=1}^T |\mathbf{W}_{T,t}(\tau)|\,E\left(|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t|^2\,|\mathscr{F}_{t-1}\right)\right| = O_P(1)$.

Finally, I turn to $J_{T,3}$, and apply the truncation method. Let $\mathbf{u}_t = \boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t$, $\mathbf{u}_t' = \mathbf{u}_t I\left(\|\mathbf{u}_t\| \leq T^{\frac{2}{\delta}}\right)$

and $\mathbf{u}_t'' = \mathbf{u}_t - \mathbf{u}_t'$. Then I have

$$\begin{aligned} J_{T,3} &= \max_{1\le l\le N_T}\left|\sum_{t=1}^T (\mathbf{I}_d\otimes \mathbf{W}_{T,t}(s_l))\left(\mathbf{u}_t'+\mathbf{u}_t''-E(\mathbf{u}_t'+\mathbf{u}_t''|\mathscr{F}_{t-1})\right)\right| \\ &\le \max_{1\le l\le N_T}\left|\sum_{t=1}^T (\mathbf{I}_d\otimes \mathbf{W}_{T,t}(s_l))\left(\mathbf{u}_t'-E(\mathbf{u}_t'|\mathscr{F}_{t-1})\right)\right| + \max_{1\le l\le N_T}\left|\sum_{t=1}^T (\mathbf{I}_d\otimes \mathbf{W}_{T,t}(s_l))\,\mathbf{u}_t''\right| \\ &\quad + \max_{1\le l\le N_T}\left|\sum_{t=1}^T (\mathbf{I}_d\otimes \mathbf{W}_{T,t}(s_l))\,E(\mathbf{u}_t''|\mathscr{F}_{t-1})\right| = J_{T,31}+J_{T,32}+J_{T,33}. \end{aligned}$$

It's easy to show that $J_{T,32} = O_P(T^{\frac{2}{\delta}}d_T)$ and $J_{T,33} = O_P(T^{\frac{2}{\delta}}d_T)$. Thus, I focus on $J_{T,31}$.

For any $1\le l\le N_T$, let $\mathbf{Y}_t = (\mathbf{I}_d\otimes \mathbf{W}_{T,t}(s_l))(\mathbf{u}_t' - E(\mathbf{u}_t'|\mathscr{F}_{t-1}))$, then I have $E\left(\mathbf{Y}_t|\mathscr{F}_{t-1}\right)=0$ and $|\mathbf{Y}_t|\le 2T^{2/\delta}d_T$. Also,

$$\max_{1\le l\le N_T}\sum_{t=1}^T |\mathbf{W}_{T,t}(s_l)|\,E\left(|\boldsymbol{\zeta}_t\boldsymbol{\varepsilon}_t|^2\,|\mathscr{F}_{t-1}\right) = O_P(1),$$

which yields

$$\max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| \le Md_T\max_{1\le l\le N_T}\sum_{t=1}^T|\mathbf{W}_{T,t}(s_l)|\,E\left(|\mathbf{u}_t|^2\,|\mathscr{F}_{t-1}\right) = O_P\left(d_T\right).$$

Therefore, I have $\max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right| = O_P(d_T)$. By Lemma A.1.2 and choosing $\beta=4$, I have

$$\begin{aligned} \Pr\left(J_{T,31}>\sqrt{\beta M}\gamma_T\right) &= \Pr\left(J_{T,31}>\sqrt{\beta M}\gamma_T,\ \max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|\le Md_T\right) \\ &+\Pr\left(J_{T,31}>\sqrt{\beta M}\gamma_T,\ \max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|> Md_T\right) \\ &\le \Pr\left(J_{T,31}>\sqrt{\beta M}\gamma_T,\ \max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|\le Md_T\right) \\ &+\Pr\left(\max_{1\le l\le N_T}\left|\sum_{t=1}^T E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|> Md_T\right) \le N_T\exp\left(-\frac{\beta M\gamma_T^2}{2(Md_T+M\gamma_T T^{\frac{2}{\delta}}d_T)}\right)+o(1) \\ &\le N_T\exp\left(-\frac{\beta}{2}\log T\right) = N_T T^{-\frac{\beta}{2}} = o(1) \end{aligned}$$

given $d_T T^{\frac{4}{\delta}}\log T\to 0$.

I now have completed the proof of the second result. □

Proof of Lemma A.1.6.

Let $\boldsymbol{\Psi}_j(\tau) = \mathbf{J}\boldsymbol{\Phi}^j(\tau)\mathbf{J}^\top$, where

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

and $\mathbf{J} = \left[\mathbf{I}_d, \mathbf{0}_{d\times d(p-1)}\right]$.

To proceed, I write $\mathbf{y}_t$ as a time–varying VMA(∞):

$$\boldsymbol{y_t} = \sum_{j=0}^{\infty} \mathbf{\Psi}_{j,t} \left(\sum_{l=0}^{q} \mathbf{B}_l(\tau_{t-j})\mathbf{x}_{t-l-j} + \boldsymbol{\eta}_{t-j} \right) = \boldsymbol{\mu}_t + \sum_{j=0}^{\infty} \mathbf{D}_{j,t}^{\boldsymbol{\varepsilon}} \boldsymbol{\varepsilon}_{t-j} + \sum_{l=0}^{q}\sum_{j=0}^{\infty} \mathbf{D}_{j,l,t}^{\mathbf{v}} \mathbf{v}_{t-l-j},$$

where $\boldsymbol{\mu}_t = \sum_{j=0}^{\infty}\sum_{l=0}^{q} \mathbf{\Psi}_{j,t}\mathbf{B}_l(\tau_{t-j})\mathbf{g}(\tau_{t-l-j})$, $\mathbf{\Psi}_{j,t} = \mathbf{J}\prod_{m=0}^{j-1}\mathbf{\Phi}(\tau_{t-m})\mathbf{J}^\top$, $\mathbf{D}_{j,t}^{\boldsymbol{\varepsilon}} = \mathbf{\Psi}_{j,t}\boldsymbol{\omega}(\tau_{t-j})$, and $\mathbf{D}_{j,l,t}^{\mathbf{v}} = \sum_{k=0}^{j} \mathbf{\Psi}_{k,t}\mathbf{B}_l(\tau_{t-k})\mathbf{C}_{j-k}(\tau_{t-l-k})$.

Let ρ denote the largest eigenvalue of $\mathbf{\Phi}(\tau)$ uniformly over $\tau \in [0,1]$. Then I have $\rho < 1$ by Assumption 2.2.1.1. In addition, similar to the proof of Proposition 2.4 in Dahlhaus and Polonik [2009], I have $\max_{t\geq 1} |\prod_{m=0}^{j-1}\mathbf{\Phi}(\tau_{t-m})| \leq M\rho^j$.

Next, I will show that $\mathbf{y}_t$ can be approximated by a time-varying MA(∞) process $\widetilde{\mathbf{y}}_t$ satisfying $\|\mathbf{y}_t - \widetilde{\mathbf{y}}_t\|_\delta = O(T^{-1})$, where $\widetilde{\mathbf{y}}_t$ has been defined in the body of this lemma. It follows that

$$\begin{aligned} \{E\|\mathbf{y}_t - \widetilde{\mathbf{y}}_t\|^\delta\}^{1/\delta} &\leq M\left(|\boldsymbol{\mu}_t - \boldsymbol{\mu}(\tau_t)| + \sum_{j=0}^{\infty} \left|\mathbf{D}_{j,t}^{\boldsymbol{\varepsilon}} - \mathbf{D}_j^{\boldsymbol{\varepsilon}}(\tau_t)\right| + \sum_{l=0}^{q}\sum_{j=0}^{\infty} \left|\mathbf{D}_{j,l,t}^{\mathbf{v}} - \mathbf{D}_{j,l}^{\mathbf{v}}(\tau_t)\right| \right) \\ &:= O(1)\cdot (I_{T,1} + I_{T,2} + I_{T,3}), \end{aligned}$$

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

Consider $I_{T,1}$. Note that for any conformable matrices $\{\mathbf{A}_i\}$ and $\{\mathbf{B}_i\}$, since

$$\prod_{i=1}^{r}\mathbf{A}_i - \prod_{i=1}^{r}\mathbf{B}_i = \sum_{j=1}^{r}\left(\prod_{k=1}^{j-1}\mathbf{A}_k\right)(\mathbf{A}_j - \mathbf{B}_j)\left(\prod_{k=j+1}^{r}\mathbf{B}_k\right),$$

I obtain

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

Hence, I have

$$\begin{aligned} I_{T,1} &\leq \sum_{j=0}^{\infty}|\mathbf{\Psi}_{j,t} - \mathbf{\Psi}_j(\tau_t)|\cdot\sum_{l=0}^{q}|\mathbf{B}_l(\tau_{t-j})\mathbf{g}(\tau_{t-l-j})| \\ &\quad + \sum_{j=0}^{\infty}|\mathbf{\Psi}_j(\tau_t)|\cdot\sum_{l=0}^{q}|\mathbf{B}_l(\tau_{t-j})\mathbf{g}(\tau_{t-l-j}) - \mathbf{B}_l(\tau_t)\mathbf{g}(\tau_t)| \\ &\leq M\sum_{j=0}^{\infty}\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1} + M\sum_{j=0}^{\infty}\rho^j\frac{j}{T} = O(T^{-1}), \end{aligned}$$

where I have used the facts that $|\mathbf{\Psi}_j(\tau)| \leq M\rho^j$ and

$$\begin{aligned} & |\mathbf{B}_l(\tau_{t-j})\mathbf{g}(\tau_{t-l-j}) - \mathbf{B}_l(\tau_t)\mathbf{g}(\tau_t)| \\ = \; & |\mathbf{B}_l(\tau_{t-j})\mathbf{g}(\tau_{t-l-j}) - \mathbf{B}_l(\tau_t)\mathbf{g}(\tau_{t-l-j}) + \mathbf{B}_l(\tau_t)\mathbf{g}(\tau_{t-l-j}) - \mathbf{B}_l(\tau_t)\mathbf{g}(\tau_t)| \\ \leq \; & |\mathbf{B}_l(\tau_{t-j}) - \mathbf{B}_l(\tau_t)|\cdot|\mathbf{g}(\tau_{t-l-j})| + |\mathbf{B}_l(\tau_t)|\cdot|\mathbf{g}(\tau_{t-l-j}) - \mathbf{g}(\tau_t)| \leq M\frac{j}{T}. \end{aligned}$$

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

For $I_{T,3}$,

$$
\begin{aligned}
I_{T,3} &\leq \sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{j} |\mathbf{\Psi}_{k,t}\mathbf{B}_l(\tau_{t-k})\mathbf{C}_{j-k}(\tau_{t-l-k}) - \mathbf{\Psi}_k(\tau_t)\mathbf{B}_l(\tau_t)\mathbf{C}_{j-k}(\tau_t)| \\
&= \sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} |\mathbf{\Psi}_{j,t}\mathbf{B}_l(\tau_{t-j})\mathbf{C}_k(\tau_{t-l-j}) - \mathbf{\Psi}_j(\tau_t)\mathbf{B}_l(\tau_t)\mathbf{C}_k(\tau_t)| \\
&\leq \sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} |\mathbf{\Psi}_{j,t} - \mathbf{\Psi}_j(\tau_t)| \cdot |\mathbf{B}_l(\tau_{t-j})| \cdot |\mathbf{C}_k(\tau_{t-l-j})| \\
&\quad + \sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} |\mathbf{\Psi}_j(\tau_t)| \cdot |\mathbf{B}_l(\tau_{t-j}) - \mathbf{B}_l(\tau_t)| \cdot |\mathbf{C}_k(\tau_{t-l-j})| \\
&\quad + \sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} |\mathbf{\Psi}_j(\tau_t)| \cdot |\mathbf{B}_l(\tau_t)| \cdot |\mathbf{C}_k(\tau_{t-l-j}) - \mathbf{C}_k(\tau_t)| \\
&\leq M\sum_{j=0}^{\infty}\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1} \cdot \sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}|\mathbf{C}_k(\tau)| \\
&\quad + \frac{M}{T}\left(\sum_{j=0}^{\infty} j\rho^j\right)\left(\sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}|\mathbf{C}_k(\tau)| + \sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}\left|\mathbf{C}_k^{(1)}(\tau)\right|\right) = O(T^{-1}).
\end{aligned}
$$

In addition, it is straightforward to verify that

$$
\sup_{\tau\in[0,1]}\sum_{j=0}^{\infty} j|\mathbf{D}_j^{\boldsymbol{\varepsilon},(k)}(\tau)| < \infty \quad \text{and} \quad \sup_{\tau\in[0,1]}\sum_{j=0}^{\infty} j|\mathbf{D}_{j,l}^{\mathbf{v},(k)}(\tau)| < \infty
$$

for $k = 0, 1$ (see, the proof of Propositions 2.2.1, for example.)

The proof is now completed. □

Proof of Lemma A.1.7.

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

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

for $k = 0, 1, 2$. Since $\mathbf{Z}_t$ can be approximated by a time-varying VMA(∞) process by Lemma A.1.6, then the uniform convergence results follows directly from Lemma 2.2.3.

(3). Part (3) follows directly from Lemma A.1.5.3.

□

Proof of Lemma A.1.8.

(1). By Lemma A.1.7,

$$\mathbf{S}\mathbf{X}_{\mathbf{C}} = \begin{pmatrix} (\mathbf{X}_{\widetilde{\mathbf{C}},1}^\top, \mathbf{0}_{d\times(l-s)})(\mathbf{X}_{\widetilde{\mathbf{C}},\tau_1}^\top \mathbf{K}_{\tau_1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau_1})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau_1}^\top \mathbf{K}_{\tau_1}\mathbf{X}_{\mathbf{C}} \\ \vdots \\ (\mathbf{X}_{\widetilde{\mathbf{C}},T}^\top, \mathbf{0}_{d\times(l-s)})(\mathbf{X}_{\widetilde{\mathbf{C}},\tau_T}^\top \mathbf{K}_{\tau_T}\mathbf{X}_{\widetilde{\mathbf{C}},\tau_T})^{-1}\mathbf{X}_{\widetilde{\mathbf{C}},\tau_T}^\top \mathbf{K}_{\tau_T}\mathbf{X}_{\mathbf{C}} \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{X}_{\widetilde{\mathbf{C}},1}^\top \boldsymbol{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}^{-1}(\tau_1)\boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}^\top(\tau_1) \\ \vdots \\ \mathbf{X}_{\widetilde{\mathbf{C}},T}^\top \boldsymbol{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}^{-1}(\tau_T)\boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}^\top(\tau_T) \end{pmatrix}(1+o_P(1)),$$

which follows that

$$\frac{1}{T}\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT}-\mathbf{S})^\top(\mathbf{I}_{dT}-\mathbf{S})\mathbf{X}_{\mathbf{C}}$$
$$= \frac{1}{T}\sum_{t=1}^T \left(\mathbf{X}_{\mathbf{C},t}^\top - \mathbf{X}_{\widetilde{\mathbf{C}},t}^\top \boldsymbol{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}^{-1}(\tau_t)\boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}^\top(\tau_t)\right)^\top \left(\mathbf{X}_{\mathbf{C},t}^\top - \mathbf{X}_{\widetilde{\mathbf{C}},t}^\top \boldsymbol{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}^{-1}(\tau_t)\boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}^\top(\tau_t)\right) + o_P(1).$$

Note that each element of $\boldsymbol{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}(\tau)$ and $\boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau)$ is Lipschitz continuous. Thus, by Lemmas 2.2.2 and A.1.6, the result holds.

(2). Let $\rho_T = h^2 + \sqrt{\log T/(Th)}$. By Lemma A.1.7.1, I have

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

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

$$\mathbf{X}_{\mathbf{C}}^\top(\mathbf{I}_{dT}-\mathbf{S})^\top(\mathbf{I}_{dT}-\mathbf{S})\widetilde{\mathbf{X}}$$
$$= \sum_{t=1}^T \left(\mathbf{X}_{\mathbf{C},t}\mathbf{X}_{\widetilde{\mathbf{C}},t}^\top - \boldsymbol{\Sigma}_{\mathbf{X}_{\mathbf{C},\widetilde{\mathbf{C}}}}(\tau_t)\boldsymbol{\Sigma}_{\mathbf{X}_{\widetilde{\mathbf{C}}}}^{-1}(\tau_t)\mathbf{X}_{\widetilde{\mathbf{C}},t}\mathbf{X}_{\widetilde{\mathbf{C}},t}^\top(1+O_P(\rho_T))\right)\boldsymbol{\theta}(\tau_t)\cdot O_P(\rho_T) = O_P(T\rho_T^2),$$

where the last equality follows from Lemma 2.2.2. Finally, the result holds since $O_P(T\rho_T^2) = o_P(\sqrt{T})$ by Assumption 2.4.2. □

Appendix B

Appendix for Chapter 3

For the sake of presentation, I first summarize the relevant mathematical symbols in Appendix B.1. Appendix B.2 includes the preliminary lemmas. Appendix B.3 presents the proofs of the main results of the paper, while Appendix B.4 provides the proofs of the preliminary lemmas.

B.1 Mathematical Symbols

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

$$\mathbf{\Sigma}(\tau) = \begin{bmatrix} 1 & \boldsymbol{\mu}^\top(\tau) & \cdots & \boldsymbol{\mu}^\top(\tau) \\ \boldsymbol{\mu}(\tau) & \mathbf{\Sigma}_0(\tau) & \cdots & \mathbf{\Sigma}_{p-1}^\top(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu}(\tau) & \mathbf{\Sigma}_{p-1}(\tau) & \cdots & \mathbf{\Sigma}_0(\tau) \end{bmatrix}, \tag{B.1.1}$$

in which $\boldsymbol{\mu}(\tau)$ and $\mathbf{B}_j(\tau)$ are defined in Proposition 3.2.1 and $\mathbf{\Sigma}_m(\tau) = \boldsymbol{\mu}(\tau)\boldsymbol{\mu}(\tau)^\top + \sum_{j=0}^{\infty} \mathbf{B}_j(\tau)\mathbf{B}_{j+m}^\top(\tau)$ for $m = 0, \ldots, p-1$. I define the estimator of $\mathbf{\Sigma}(\tau)$ as

$$\widehat{\mathbf{\Sigma}}(\tau) = \left(\frac{1}{T}\sum_{t=1}^{T} K_h(\tau_t - \tau)\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}^\top K_h(\tau_t - \tau), \tag{B.1.2}$$

where $\mathbf{z}_t$ is defined in (3.2.3).

Next, I let

$$\mathbf{V}(\tau) = \begin{bmatrix} \mathbf{V}_{1,1}(\tau) & \mathbf{V}_{2,1}^\top(\tau) \\ \mathbf{V}_{2,1}(\tau) & \mathbf{V}_{2,2}(\tau) \end{bmatrix}, \tag{B.1.3}$$

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

$$\begin{aligned} \mathbf{V}_{2,1}(\tau) &= \lim_{T\to\infty} \frac{h}{T}\sum_{t=1}^{T} E\left(\mathrm{vech}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top)\boldsymbol{\eta}_t^\top \mathbf{Z}_{t-1}^\top\right) K_h(\tau_t - \tau)^2 \cdot (\mathbf{\Sigma}^{-1}(\tau) \otimes \mathbf{I}_d), \\ \mathbf{V}_{2,2}(\tau) &= \lim_{T\to\infty} \frac{h}{T}\sum_{t=1}^{T} E\left(\mathrm{vech}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top)\mathrm{vech}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top)^\top\right) K_h(\tau_t - \tau)^2 \\ &\quad - \tilde{v}_0 \mathrm{vech}\left(\mathbf{\Omega}(\tau)\right) \mathrm{vech}\left(\mathbf{\Omega}(\tau)\right)^\top. \end{aligned} \tag{B.1.4}$$

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

$$\widehat{\mathbf{V}}(\tau) = \begin{bmatrix} \widehat{\mathbf{V}}_{1,1}(\tau) & \widehat{\mathbf{V}}_{2,1}^\top(\tau) \\ \widehat{\mathbf{V}}_{2,1}(\tau) & \widehat{\mathbf{V}}_{2,2}(\tau) \end{bmatrix}, \tag{B.1.5}$$

where $\widehat{\mathbf{V}}_{1,1}(\tau)$, $\widehat{\mathbf{V}}_{2,1}(\tau)$ and $\widehat{\mathbf{V}}_{2,2}(\tau)$ have the forms identical to their counterparts of (B.1.4), but I replace $\mathbf{\Sigma}(\tau)$, $\boldsymbol{\eta}_t$ and $\mathbf{\Omega}(\tau)$ with their estimators presented in (B.1.2) and (3.2.4).

Recall the following definition:

$$\mathbf{\Phi}(\tau) = \begin{bmatrix} \mathbf{A}_1(\tau) & \cdots & \mathbf{A}_{p-1}(\tau) & \mathbf{A}_p(\tau) \\ \mathbf{I}_d & \cdots & \mathbf{0}_d & \mathbf{0}_d \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_d & \cdots & \mathbf{I}_d & \mathbf{0}_d \end{bmatrix}. \tag{B.1.6}$$

Replacing $\mathbf{A}_j(\tau)$'s of (B.1.6) with their estimators obtained from (3.2.4) yields an estimator, $\widehat{\mathbf{\Phi}}(\tau)$, for $\mathbf{\Phi}(\tau)$.

Finally, I define

$$\mathbf{V}_{2,2}^*(\tau) = (\Sigma^{*,-2}(\tau)\Delta^*(\tau)) \otimes \mathbf{\Omega}^*(\tau), \tag{B.1.7}$$

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

$$\Sigma_{\pi x_1}(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\pi_t x_{1,t})K_h(\tau_t-\tau).$$

B.2 Preliminary Lemmas

Lemma B.2.1. Suppose $\{Z_t, \mathscr{F}_t\}$ 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.2.1 can be found in Hall and Heyde [1980].

Lemma B.2.2. Let $\{Z_t, \mathscr{F}_t\}$ be a martingale difference sequence. Suppose that $|Z_t| \le M$ for a constant M, $t = 1, \dots, T$. Let $V_T = \sum_{t=1}^T \mathrm{Var}(Z_t|\mathscr{F}_{t-1}) \le V$ for some $V > 0$. Then for any given $\nu > 0$,

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

Lemma B.2.2 is the Proposition 2.1 of Freedman [1975].

Assumption B.2.1. $\max_t|\boldsymbol{\mu}_t| < \infty$, $\max_t\sum_{j=1}^\infty j|\mathbf{B}_{j,t}| < \infty$, $\limsup_{T\to\infty}\sum_{t=1}^{T-1}|\boldsymbol{\mu}_{t+1}-\boldsymbol{\mu}_t| < \infty$ and $\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^\infty j|\mathbf{B}_{j,t+1}-\mathbf{B}_{j,t}| < \infty$. Define the stochastic process of the form $\mathbf{h}_t = \boldsymbol{\mu}_t + \sum_{j=0}^\infty \mathbf{B}_{j,t}\boldsymbol{\varepsilon}_t$ for $t = 1, ..., T$.

Lemma B.2.3. Let Assumptions 3.2.2 and B.2.1 hold and $\max_{t\ge1} E\left(|\boldsymbol{\varepsilon}_t|^4 |\mathscr{F}_{t-1}\right) < \infty$ a.s.. In addition, let $\{\mathbf{W}_{T,t}(\cdot)\}_{t=1}^T$ be a sequence of $q\times d$ matrices of deterministic functions, in which q is fi xed, each functional component is Lipschitz continuous and defined on a compact set $[a,b]$. Moreover, suppose that

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

Then as $T\to\infty$,

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

Lemma B.2.3 is the Lemma 2.2.3 of Chapter 2.

Lemma B.2.4. Let Assumptions 3.2.2 and B.2.1 hold. In addition, let $\{\mathbf{W}_{T,t}(\cdot)\}_{t=1}^{T}$ be a sequence of $q\times d$ matrices of deterministic functions, in which q is fixed, each functional component is Lipschitz continuous and defined on a compact set $[a,b]$. Moreover, suppose that

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

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

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

Lemma B.2.4 is the Lemma 2.2.2 of Chapter 2.

Lemma B.2.5. Suppose Assumptions 3.2.1–3.2.3 hold. Let $\mathbf{W}(\cdot)$ be a twice-differentiable functional matrix in $R^{m\times d}$. As $T\to\infty$,

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

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

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

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

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

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

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

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

Lemma B.2.6. Let Assumptions 3.2.1–3.2.3 hold. Suppose $\max_{t\geq 1} E[|\boldsymbol{\varepsilon}_t|^4 | \mathscr{F}_{t-1}] < \infty$ a.s. and $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$. As $T \to \infty$,

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

Lemma B.2.7. Let Assumptions 3.2.1–3.2.3 hold. Suppose $\max_{t\geq 1} E[|\boldsymbol{\varepsilon}_t|^4 | \mathscr{F}_{t-1}] < \infty$ a.s. and $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$. As $T \to \infty$,

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

Lemma B.2.8. Let Assumptions 3.2.1–3.3.1 hold. Suppose $\max_{t\geq 1} E[|\boldsymbol{\varepsilon}_t|^4 | \mathscr{F}_{t-1}] < \infty$ a.s. and $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$. Then,

1. $\sup_{\tau\in[h,1-h]} \left|\mathbf{s}(\tau)\mathbf{X}_\tau - \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau)\boldsymbol{\Sigma}_{\mathbf{ZX}_1}(\tau)\otimes[1,0]\right| = O_P(\rho_T)$, where $\boldsymbol{\Sigma}_{\mathbf{Z}}(\tau) = \boldsymbol{\Sigma}(\tau)\otimes\mathbf{I}_d$, $\boldsymbol{\Sigma}_{\mathbf{ZX}_1}(\tau) = \boldsymbol{\Sigma}_{\mathbf{z}x_1}\otimes\mathbf{I}_d$ and $\boldsymbol{\Sigma}_{\mathbf{z}x_1} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\mathbf{z}_{t-1}x_{1,t})K_h(\tau_t-\tau)$;
2. $\sup_{\tau\in[h,1-h]} \left|\mathbf{s}(\tau)\mathbf{W}_\tau - \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau)\boldsymbol{\Sigma}_{\mathbf{ZW}}(\tau)\otimes[1,0]\right| = O_P(\rho_T)$, where $\boldsymbol{\Sigma}_{\mathbf{ZW}}(\tau) = \boldsymbol{\Sigma}_{\mathbf{z}\pi}(\tau)\otimes\mathbf{I}_d$ and $\boldsymbol{\Sigma}_{\mathbf{z}\pi}(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\mathbf{z}_{t-1}\pi_t)K_h(\tau_t-\tau)$;
3. $\sup_{\tau\in[h,1-h]} \left|\frac{1}{T}\mathbf{W}_\tau^\top(\mathbf{I}_{dT}-\mathbf{S})^\top\mathbf{K}_\tau(\mathbf{I}_{dT}-\mathbf{S})\mathbf{X}_\tau - (\Sigma^*(\tau)\otimes\mathbf{I}_d)\otimes\boldsymbol{\Lambda}_1\right| = O_P(\rho_T)$, where $\Sigma^*(\tau) = \Sigma_{\pi x_1}(\tau) - \boldsymbol{\Sigma}_{\mathbf{z}\pi}^\top(\tau)\boldsymbol{\Sigma}^{-1}(\tau)\boldsymbol{\Sigma}_{\mathbf{z}x_1}(\tau)$, $\Sigma_{\pi x_1}(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\pi_t x_{1,t})K_h(\tau_t-\tau)$ and $\boldsymbol{\Lambda}_1 = \text{diag}(1,\widetilde{c}_2)$;
4. if $Th^9 \to 0$,
$$\mathbf{W}_\tau^\top(\mathbf{I}_{dT}-\mathbf{S})^\top\mathbf{K}_\tau(\mathbf{I}_{dT}-\mathbf{S})\begin{bmatrix}\mathbf{Z}_0^\top\text{vec}(\mathbf{A}^*(\tau_1))\\ \vdots \\ \mathbf{Z}_{T-1}^\top\text{vec}(\mathbf{A}^*(\tau_T))\end{bmatrix} = o_P(\sqrt{Th});$$
5. if $Th^9 \to 0$,
$$\frac{1}{\sqrt{Th}}\sum_{t=1}^T[\mathbf{W}_t - \boldsymbol{\Sigma}_{\mathbf{ZW}}^\top(\tau_t)\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau_t)\mathbf{Z}_{t-1}(1+O_P(\rho_T))]$$
$$\times\mathbf{Z}_{t-1}^\top\mathbf{s}(\tau_t)\begin{bmatrix}\mathbf{X}_{1,1}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_1)-\boldsymbol{\omega}_{\cdot,1}^*(\tau_t))\\ \vdots \\ \mathbf{X}_{1,T}(\boldsymbol{\omega}_{\cdot,1}^*(\tau_T)-\boldsymbol{\omega}_{\cdot,1}^*(\tau_t))\end{bmatrix}K_h(\tau_t-\tau) = o_P(1).$$

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.2.9. Suppose Assumption 3.2.3 hold. Then, I 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. $\max_{2\leq t\leq T} a_t = O(1/T)$;

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

4. $T\sum_{s=1}^{T-1} b_s^2 = O(1)$;

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

Lemma B.2.10. Let $\boldsymbol{\xi}_1$,...,$\boldsymbol{\xi}_T$ be a d-dimensional martingale difference for which $\boldsymbol{\xi}_t \in \mathscr{L}^p$, $p > 1$. Let $p^* = \min(2, p)$. Then

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

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

Lemma B.2.11. Suppose Assumption 3.2.3 hold. Assume $\mathbf{w}_t^* \in \mathscr{L}^{\delta/2}$ for some $\delta > 4$ is $\mathscr{F}_t$-measurable. Then, as $T \to \infty$,

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

Lemma B.2.12. Suppose Assumptions 3.2.1–3.2.3 hold and $\mathbf{y}_t = \mathbf{Z}_{t-1}\boldsymbol{\eta}_t$. Then, as $T \to \infty$,

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

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

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

Lemma B.2.13. Suppose Assumptions 3.2.1–3.2.3 hold and $\mathbf{y}_t = \mathbf{Z}_{t-1}\boldsymbol{\eta}_t$. Then as $T \to \infty$,

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

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

Lemma B.2.14. Let Assumptions 3.2.1–3.2.3 hold. Suppose further that $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$, $\max_{t\geq 1} E[|\boldsymbol{\varepsilon}_t|^4|\mathscr{F}_{t-1}] < \infty$ a.s., each element of $\mathbf{A}(\cdot)$ has finite third-order derivative, $Th^6 \to 0$, and $Th^2/(\log T)^2 \to \infty$. As $T \to \infty$, I have

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

where $\mathbf{V}_{\boldsymbol{\beta}}(\tau) := \boldsymbol{\Sigma}^{-1}(\tau)\otimes\boldsymbol{\Omega}(\tau)$ and $\boldsymbol{\Sigma}(\tau)$ is defined in (B.1.1).

B.3 Proofs of the Main Results

Proof of Proposition 3.2.1.

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

First, I investigate the validity of the VMA representations of $\mathbf{x}_t$ and $\widetilde{\mathbf{x}}_t$. Let ρ_A denote the largest eigenvalue of $\boldsymbol{\Phi}(\tau)$ uniformly over $\tau \in [0,1]$. Then, $\rho_A < 1$ by Assumption 3.2.1. Similar to the proof of Proposition 2.4 in Dahlhaus and Polonik [2009], I 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 $|\mathrm{E}(\boldsymbol{x_t})| \leq \sum_{j=0}^{\infty}|\boldsymbol{\Psi}_{j,t}|\cdot|\mathbf{a}(\tau_{t-j})| \leq M\sum_{j=0}^{\infty}\rho_A^j < \infty$ and

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

Similarly, I have $|\mathrm{E}\left(\widetilde{\mathbf{x}}_t\right)| < \infty$ and $|\mathrm{Var}\left(\widetilde{\mathbf{x}}_t\right)| < \infty$.

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

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

which implies for the same $\delta > 4$ as in Assumption 3.2.2,

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

The proof is now completed. □

Proof of Theorem 3.2.1.

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

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

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

I now begin our investigation. Since

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

I write

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

By standard arguments for the local linear kernel estimator and the uniform convergence results in Lemmas B.2.5.2-3, I 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.2.6.1, I 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). I begin our investigation on the asymptotic normality by writing that for $\forall\tau\in(0,1)$,

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

Let $\rho_T = h^2 + \sqrt{\frac{\log T}{Th}}$. By what I have just proved for Theorem 3.2.1 (i), for $\forall \tau \in [0,1]$ I have

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

By Lemma B.2.6.2, $\mathbf{I}_{T,6}$ and $\mathbf{I}_{T,7}$ are both $o_P((Th)^{-1/2})$. Hence,

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

Combined with the convergence results of the sample covariance matrix stated in Lemma B.2.5, the above development yields that

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

Below, I focus on $\mathbf{I}_{T,8}$. First, I show $\text{Var}(\mathbf{I}_{T,8}) \to \mathbf{V}(\tau)$. Let

$$\text{Var}(\mathbf{I}_{T,8}) = \begin{bmatrix}\widetilde{\mathbf{V}}_{1,1}(\tau) & \widetilde{\mathbf{V}}_{2,1}^\top(\tau) \\ \widetilde{\mathbf{V}}_{2,1}(\tau) & \widetilde{\mathbf{V}}_{2,2}(\tau)\end{bmatrix},$$

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

By construction and Assumption 3.2.2, $\mathbf{I}_{T,8}$ is a summation of m.d.s., I thus use Lemma B.2.1 and Cramér-Wold device to prove its asymptotic normality. It suffices to show that $\mathbf{d}^\top\mathbf{I}_{T,8} \to_D N\left(\mathbf{0}, \mathbf{d}^\top\mathbf{V}(\tau)\mathbf{d}\right)$ for any conformable unit vector $\mathbf{d}$. Let

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

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

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

Since conditional on $\mathscr{F}_{t-1}$ the third and fourth moments of $\boldsymbol{\varepsilon}_t$ are identical to the corresponding unconditional moments a.s., by Lemma B.2.5.1 I can prove that $\sum_{t=1}^T E(\mathbf{Z}_{T,t}^2(\tau)|\mathscr{F}_{t-1}) \to_P \mathbf{d}^\top\mathbf{V}(\tau)\mathbf{d}$.

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

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

Thus, the CLT follows.

Finally, I consider $\widehat{\mathbf{V}}(\cdot)$. By Lemma B.2.5 and the above proof, I have $\widehat{\mathbf{V}}_{1,1}(\tau) \to_P \mathbf{V}_{1,1}(\tau)$. By the uniform convergence results of $\widehat{\mathbf{A}}(\tau)$, I can replace $\widehat{\boldsymbol{\eta}}_t$ with $\boldsymbol{\eta}_t$ in the following derivations. Therefore, I have

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

and

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

The proof is now completed. □

Proof of Theorem 3.2.2.

I 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 $\mathsf{p}\le\mathsf{P}$.

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

(ii) I then consider $\mathsf{p}>p$. Lemma B.2.7 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 $(\mathsf{p}-p)\chi_T\ge\chi_T$, which converges to zero at a slower rate than ρ_T^2, it follows that

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

The proof is now completed. □

Proof of Theorem 3.2.3.

First, I introduce some additional notation to facilitate the development. Let $A_Q=\widetilde{v}_0\cdot\text{tr}\left\{\int_0^1\boldsymbol{\Sigma}_Q(\tau)\text{d}\tau\right\}$ and $B_Q=4C_B\cdot\text{tr}\left\{\int_0^1\boldsymbol{\Sigma}_Q(\tau)^2\text{d}\tau\right\}$, where

$$\boldsymbol{\Sigma}_Q(\tau)=\mathbf{H}(\tau)^{1/2}\mathbf{C}\mathbf{V}_{\boldsymbol{\beta}}(\tau)\mathbf{C}^\top\mathbf{H}(\tau)^{1/2}.$$

Recall $\rho_T=h^2+\sqrt{\frac{\log T}{Th}}$. By Theorem 3.2.1.1, I have

$$\sup_{\tau\in[0,1]}\left|\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right|=O_P\left(\rho_T\right). \tag{B.3.1}$$

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

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

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

In addition, by Lemma B.2.5 and Lemma B.2.6.1, I have

$$\sup_{[h,1-h]} |\mathbf{S}_T(\tau) - \mathbf{\Sigma}_{\mathbf{Z}}(\tau)\otimes \mathbf{\Lambda}_1| = O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right),$$
$$\sup_{[0,1]} |\mathbf{R}_T(\tau)| = O_P\left(\sqrt{\frac{\log T}{Th}}\right), \tag{B.3.2}$$

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

$$\sup_{[h,1-h]}\left|\mathbf{C}\widehat{\boldsymbol{\beta}}(\tau) - \mathbf{c} - \mathbf{C}\mathbf{\Sigma}_{\mathbf{Z}}^{-1}(\tau)\mathbf{R}_T(\tau)\right| = O_P(\rho_T^2). \tag{B.3.3}$$

By (B.3.2) and $Th^{11/2}\to 0$, I have

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

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

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

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

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

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

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

Then I have

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

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

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

I have

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

where the definition of $\widetilde{U}$ is obvious. By Lemma B.2.14, I have

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

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

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

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

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

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

Combining the above results, I have proved

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

Note that

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

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

Since $\widehat{\mathbf{c}}=\mathbf{c}+O_P\left(T^{-1/2}\right)$ by Lemma B.2.14, I have $I_{T,5}=O_P(T^{-1})$. For $I_{T,6}$, by (B.3.1) and (B.3.3), I have

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

provided that

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

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

Observe that

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

Finally, I need only to focus on

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

Hence, it suffices to show $T\sqrt{h}I_{T,7}=o_P(1)$. Using Lemma B.2.6.3, it is easy to know that

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

The proof is now completed. □

Proof of Corollary 3.2.1.

Under the local alternative 3.2.9, I have $\mathbf{C}\boldsymbol{\beta}(\tau) = \mathbf{c} + d_T \boldsymbol{f}(\tau)$ and thus

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

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

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

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

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

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

As $T\sqrt{h}\left(\int_0^1 \mathbf{R}_T^\top(\tau)\mathbf{H}_0(\tau)\mathbf{R}_T(\tau) - (Th)^{-1}A_Q\right) \to N(0, B_Q)$, I have

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

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

Proof of Theorem 3.3.1.

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

Let $\boldsymbol{\alpha}(\tau) = \mathrm{vec}\left(\mathbf{A}_1(\tau), ..., \mathbf{A}_p(\tau)\right)$, $\boldsymbol{\sigma}(\tau) = \mathrm{vech}\left(\boldsymbol{\Omega}(\tau)\right)$ and $\boldsymbol{\phi}(\tau) = [\boldsymbol{\alpha}^\top(\tau), \boldsymbol{\sigma}^\top(\tau)]^\top$. Given the joint distribution of $\mathrm{vec}(\widehat{\mathbf{A}}(\tau))$ and $\mathrm{vech}(\widehat{\boldsymbol{\Omega}}(\tau))$ in Theorem 3.2.1, Theorem 3.3.1 can be obtained by the Delta method. By the first-order approximation of $\mathrm{vec}\left(\widehat{\mathbf{B}}_j(\tau)\right)$ around $\mathrm{vec}\left(\mathbf{B}_j(\tau)\right)$, I have

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

and thus

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

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

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

$$\mathbf{N}_1(\tau)\triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau) = \mathbf{D}_1. \tag{B.3.4}$$

I then illustrate how to combine equation (B.3.4) with gradient equations from type (b) restrictions in order to compute $\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$.

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

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

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

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

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

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

In the case of long-run restrictions, type (b) restrictions involve $\boldsymbol{\alpha}(\tau)$, so that $\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$ has the form $[\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\omega}(\tau), \triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]$. First, equation (B.3.4) must be extended in the form $\mathbf{N}_1\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) = [\mathbf{0}, \mathbf{D}_1]$. Second, long-run restrictions can be expressed as $\mathbf{Q}\text{vec}\left[\mathbf{A}_\tau^{-1}(1)\boldsymbol{\omega}\right] = 0$, where $\mathbf{Q}$ is a $d(d-1)/2 \times d^2$ matrix of 0 and 1, and $\mathbf{A}_\tau(1) = \mathbf{I}_d - \sum_{i=1}^p \mathbf{A}_i(\tau)$. By $\mathrm{d}\mathbf{A}^{-1} = -\mathbf{A}^{-1} \cdot \mathrm{d}\mathbf{A} \cdot \mathbf{A}^{-1}$, I have

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

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

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

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

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

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

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

The proof is now completed. □

Proof of Theorem 3.3.2.

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

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

$$
\begin{aligned}
\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) &= Q_T^{-1}(\tau) \times \frac{1}{T}\sum_{t=1}^{T}(\mathbf{W}_t^* - \mathbf{W}_\tau^\top \mathbf{s}^\top(\tau_t)\mathbf{Z}_{t-1}))(\mathbf{x}_t - \mathbf{Z}_{t-1}^\top \mathbf{s}(\tau_t)\mathbf{x})K_h(\tau_t - \tau) \\
&= Q_T^{-1}(\tau) \times \frac{1}{T}\sum_{t=1}^{T}(\mathbf{W}_t^* - \mathbf{W}_\tau^\top \mathbf{s}^\top(\tau_t)\mathbf{Z}_{t-1}))(\mathbf{X}_{1,t}\boldsymbol{\omega}_{\bullet,1}^*(\tau_t) - \mathbf{Z}_{t-1}^\top \mathbf{s}(\tau_t)\begin{bmatrix}\mathbf{X}_{1,1}\boldsymbol{\omega}_{\bullet,1}^*(\tau_t) \\ \vdots \\ \mathbf{X}_{1,T}\boldsymbol{\omega}_{\bullet,1}^*(\tau_t)\end{bmatrix})K_h(\tau_t - \tau) \\
&\quad -Q_T^{-1}(\tau) \times \frac{1}{T}\sum_{t=1}^{T}(\mathbf{W}_t^* - \mathbf{W}_\tau^\top \mathbf{s}^\top(\tau_t)\mathbf{Z}_{t-1}))\mathbf{Z}_{t-1}^\top \mathbf{s}(\tau_t)\begin{bmatrix}\mathbf{X}_{1,1}(\boldsymbol{\omega}_{\bullet,1}^*(\tau_1) - \boldsymbol{\omega}_{\bullet,1}^*(\tau_t)) \\ \vdots \\ \mathbf{X}_{1,T}(\boldsymbol{\omega}_{\bullet,1}^*(\tau_T) - \boldsymbol{\omega}_{\bullet,1}^*(\tau_t))\end{bmatrix}K_h(\tau_t - \tau) \\
&\quad +Q_T^{-1}(\tau) \cdot \frac{1}{T}\sum_{t=1}^{T}(\mathbf{W}_t^* - \mathbf{W}_\tau^\top \mathbf{s}^\top(\tau_t)\mathbf{Z}_{t-1})) \\
&\quad \times \left(\mathbf{Z}_{t-1}^\top \mathrm{vec}(\mathbf{A}^*(\tau_t)) - \mathbf{Z}_{t-1}^\top \mathbf{s}(\tau_t)\begin{bmatrix}\mathbf{Z}_0^\top \mathrm{vec}(\mathbf{A}^*(\tau_1)) \\ \vdots \\ \mathbf{Z}_{T-1}^\top \mathrm{vec}(\mathbf{A}^*(\tau_T))\end{bmatrix}\right)K_h(\tau_t - \tau) \\
&\quad +Q_T^{-1}(\tau) \times \frac{1}{T}\sum_{t=1}^{T}(\mathbf{W}_t^* - \mathbf{W}_\tau^\top \mathbf{s}^\top(\tau_t)\mathbf{Z}_{t-1}))(\boldsymbol{\eta}_t^* - \mathbf{Z}_{t-1}^\top \mathbf{s}(\tau_t)\boldsymbol{\eta}^*)K_h(\tau_t - \tau) \\
&:= I_{T,1} + I_{T,2} + I_{T,3} + I_{T,4},
\end{aligned}
$$

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

By Lemma B.2.8, I 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.2.8.3, I 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 $\sup_{\tau\in[0,1]}\|\mathbf{s}(\tau)\boldsymbol{\eta}^*\| = O_P\left(h^2 + \sqrt{\frac{\log T}{Th}}\right)$ by Lemma B.2.6.1, and by similar arguments to the proof of Lemma B.2.8.5, for $\tau \in (0,1)$, I have

$$
\begin{aligned}
&\widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) - \boldsymbol{\omega}_{\bullet,1}^*(\tau) - \frac{1}{2}h^2\boldsymbol{\omega}_{\bullet,1}^{*,(2)}(\tau) \\
=\ &(\Sigma^{*,-1}(\tau) \otimes \mathbf{I}_d)\frac{1}{T}\sum_{t=1}^{T}(\mathbf{W}_t - \boldsymbol{\Sigma}_{\mathbf{ZW}}^\top(\tau_t)\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau_t)\mathbf{Z}_{t-1}))\boldsymbol{\eta}_t^* K_h(\tau_t - \tau) + o_P(1/\sqrt{Th}).
\end{aligned}
$$

Hence, I have

$$
\begin{aligned}
&\sqrt{Th}\begin{bmatrix}\mathrm{vec}\left(\widehat{\mathbf{A}}(\tau) - \mathbf{A}(\tau) - \frac{1}{2}h^2\tilde{c}_2\mathbf{A}^{(2)}(\tau)\right) \\ \widehat{\boldsymbol{\omega}}_{\bullet,1}(\tau) - \boldsymbol{\omega}_{\bullet,1}^*(\tau) - \frac{1}{2}h^2\boldsymbol{\omega}_{\bullet,1}^{*,(2)}(\tau)\end{bmatrix} \\
=\ &\begin{bmatrix}(\boldsymbol{\Sigma}^{-1}(\tau) \otimes \mathbf{I}_d)\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\mathbf{Z}_{t-1}\boldsymbol{\eta}_t K\left(\frac{\tau_t - \tau}{h}\right)\right) \\ (\Sigma^{*,-1}(\tau) \otimes \mathbf{I}_d)\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}(\mathbf{W}_t - \boldsymbol{\Sigma}_{\mathbf{ZW}}^\top(\tau_t)\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}(\tau_t)\mathbf{Z}_{t-1})\boldsymbol{\eta}_t^* K\left(\frac{\tau_t - \tau}{h}\right)\right)\end{bmatrix} + o_P(1) \\
:=\ &I_{T,5} + o_P(1).
\end{aligned}
$$

By simple calculation, I have

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

where $\mathbf{V}_{2,2}^*(\tau) = (\Sigma^{*,-2}(\tau)\Delta^*(\tau)) \otimes \boldsymbol{\Omega}^*(\tau)$, $\Delta^*(\tau) = \Sigma_\pi(\tau) - \boldsymbol{\Sigma}_{\mathbf{z}\pi}^\top(\tau)\boldsymbol{\Sigma}^{-1}(\tau)\boldsymbol{\Sigma}_{\mathbf{z}\pi}(\tau)$ 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 3.2.1, I can prove the asymptotic normality of $I_{T,5}$ by using martingale central limit theorem.

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

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

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

Similar to the proof of Theorem 3.3.1, the result is obtained by the Delta method. □

B.4 Proofs of the Preliminary Lemmas

Proof of Lemma B.2.5.

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

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

Second, by triangle inequality, Cauchy-Schwarz inequality and Proposition 3.2.1,

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

In addition, similar to the proof of Proposition 3.2.1, I have

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

and

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

Therefore, Lemma B.2.7 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.2.6.

(1). Let $\{S_l\}$ be a finite number of sub-intervals covering the interval $[0,1]$, which are centered at s_l with the length $\delta_T = o(h^2)$. Denote the number of these intervals by N_T then $N_T = O(\delta_T^{-1})$. Hence,

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

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 I have

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

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

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

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, I have

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

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

For any fixed $1\leq l\leq N_T$, let $\mathbf{Y}_t := \frac{1}{T}(\mathbf{u}_t'-E(\mathbf{u}_t'|\mathscr{F}_{t-1}))K_h(\tau_t-s_l)$, then I have $E\left(\mathbf{Y}_t|\mathscr{F}_{t-1}\right)=0$ and $|\mathbf{Y}_t|\leq 2T^{2/\delta}d_T$ with $d_T = \max_{1\leq t\leq T}K_h(\tau_t-s_l)/T$. Also, by Lemma B.2.5, I have

$$\sup_{0\leq\tau\leq1}\frac{1}{T}\sum_{t=1}^{T}E\left(|\mathbf{Z}_{t-1}\boldsymbol{\eta}_t|^2\,|\mathscr{F}_{t-1}\right)K_h(\tau_t-\tau) = O_P(1),$$

which follows from

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

Therefore, I have $\max_{1\leq l\leq N_T}\left|\sum_{t=1}^{T}E(\mathbf{Y}_t\mathbf{Y}_t^\top|\mathscr{F}_{t-1})\right|\leq\frac{M}{Th}$ in probability. By Lemma B.2.2, I have

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

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

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

For $J_{T,1}$ to $J_{T,2}$, using Lemmas B.2.5.2-3, I 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, I ignore $\mathbf{S}_T^{-1}(\tau_t)$ and hence,

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

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

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

For $J_{T,321}$,

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

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

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

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

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

For $J_{T,3221}$,

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

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

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

Similar to the proof of $J_{T,3221}$, I can show that $J_{T,32242} = O_P\left((Th)^{-1/2}\right)$.
Let $\mathbf{w}_{t,i} = \sum_{j=0}^{\infty}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top\right) \text{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{i-1,t+i-1}^\top\right)$. For $J_{T,32241}$,

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

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

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

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

Write

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

Consider $I_{T,1}$. Similar to the proof of part (1), I 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.2.5, I have

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

Hence, by the first result of Theorem 3.2.1

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

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

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

The proof is now completed. □

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

Proof of Lemma B.2.7.

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

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

Since $\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t - E(\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t)$ is m.d.s., I 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 3.2.1.1,

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

For $I_{T,2}$,

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

By the uniform convergence results stated in Lemmas B.2.5.2–3, I 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 $\mathbf{\Sigma}(\tau) = \mathrm{plim}_{T\to\infty}\mathbf{S}_{T,0}(\tau)$, I have

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

Similar to the proof of $J_{T,4}$ in Lemma B.2.6, I can show

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

Since $(Th)^{-1}+T^{-\frac{1}{2}}h^{\frac{3}{2}}=o\left(\rho_T^2\right)$, result (1) follows.
(2). For $\mathsf{p}<p$, I have $\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau)-\boldsymbol{\Lambda}_{\mathsf{p}}(\tau)=\mathbf{B}_{\mathsf{p}}(\tau)+o_P(1)$ uniformly over $\tau\in[0,1]$, where $\mathbf{B}_{\mathsf{p}}(\tau)$ is a nonrandom bias term. Since $\widehat{\boldsymbol{\eta}}_{\mathsf{p},t}=\boldsymbol{\eta}_t+\left(\boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t)-\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t)\right)\mathbf{z}_{\mathsf{p},t-1}+\boldsymbol{\Lambda}_{\bar{\mathsf{p}}}(\tau_t)\mathbf{z}_{\bar{\mathsf{p}},t-1}$, by Lemma B.2.5.4, I have

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

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

Proof of Lemma B.2.8.

(1). By Lemma B.2.5, uniformly over $\tau\in[h,1-h]$, I have

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

(2). Similar to the proof of part (1), by Assumption 4, part (2) is easily obtained by Lemma B.2.5.
(3). By parts (1)–(2) and Lemma B.2.5, the first $d\times d$ matrix of $\frac{1}{T}\mathbf{W}_\tau^\top(\mathbf{I}_{dT}-\mathbf{S})^\top\mathbf{K}_\tau(\mathbf{I}_{dT}-\mathbf{S})\mathbf{X}_\tau$ is

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

uniformly over $\tau\in[h,1-h]$. The proofs of the other components in $\frac{1}{T}\mathbf{W}_\tau^\top(\mathbf{I}_{dT}-\mathbf{S})^\top\mathbf{K}_\tau(\mathbf{I}_{dT}-\mathbf{S})\mathbf{X}_\tau$ are similar, so omitted here.
(4). Let $\rho_T=h^2+\left(\frac{\log T}{Th}\right)^{1/2}$. By Lemma B.2.5, I have

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

uniform over $t \in [1, 2..., T]$. Hence, by Lemma B.2.5, I have

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

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

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

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

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

The proof is now completed. □

Proof of Lemma B.2.9.

(1). Write

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

(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,

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

□

Proof of Lemma B.2.10.

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

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

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 $0 < p \leq 1$, I have

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

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, I have

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

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

Proof of Lemma B.2.11.

Without loss of generality, let $E(\mathbf{w}_t^*) = \mathbf{0}$. For any integer $I \geq 1$ introduce the truncated process $\mathbf{h}_{t-1,I}^* = E\left(\mathbf{h}_{t-1}^* | \mathscr{F}_{t-I}\right)$. Then $\mathbf{h}_{t-1,I}^* = 0$ if $t \leq I$ and $\mathbf{h}_{t-1,I}^* = \sum_{s=1}^{t-I} w_{s,t}\mathbf{y}_s$ for $1 \leq I < t$. For $2 \leq t \leq T$, by Lemma B.2.10,

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

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

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

By Cauchy–Schwarz inequality, Lemma B.2.9 (iii), if $I/(Th) \to 0$, I have

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

since

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

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

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

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

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

By Lemma B.2.9 (ii), Lemma B.2.10 and Cauchy–Schwarz inequality, since $\|\mathscr{P}_{t-j}\mathbf{w}_t^*\|_{\delta/2} \leq 2\|\mathbf{w}_t^*\|_{\delta/2} < \infty$,

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

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

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

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

Proof of Lemma B.2.12.

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

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

Consider $I_{T,1}$. Write

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

Since $\mathrm{tr}\left[\left(\boldsymbol{\eta}_s\boldsymbol{\eta}_s^\top - \boldsymbol{\Omega}(\tau_s)\right)\mathbf{Z}_{s-1}^\top\mathbf{Z}_{s-1}\right]$, $s = 1, 2, \ldots$ are a martingale difference sequence and $T\sum_{t=1}^{T}w_{s,t}^2 = O(1)$ by Lemma B.2.9.2, I have $I_{T,11} = o_P(1)$. In addition, by Lemma B.2.5.4, I have $I_{T,12} = o_P(1)$.

Next, consider $I_{T,2}$. By Lemma B.2.10, Cauchy–Schwarz inequality and Lemma B.2.9.5,

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

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

Proof of Lemma B.2.14. Note that $\mathbf{y}_t^\top \mathbf{H}_t \mathbf{h}_{t-1}^*$, $t \in \mathbb{Z}$, are martingale differences with respect to the filtration $\mathscr{F}_t$. I apply Lemma B.2.1 to prove the asymptotic normality of Q_T. By $E\left(|\boldsymbol{\varepsilon}_t|^\delta |\mathscr{F}_{t-1}\right) < \infty$ a.s., I have

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

By Cauchy–Schwarz inequality, Lemma B.2.10 and Lemma B.2.9 (2), the Lindeberg condition is satisfied since

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

Apply Lemmas B.2.11 and B.2.12 with $\mathbf{w}_t^* = E\left(\mathbf{H}_t^\top \mathbf{y}_t\mathbf{y}_t^\top \mathbf{H}_t|\mathscr{F}_{t-1}\right)$, then I have the convergence of conditional variance

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

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

Proof of Lemma B.2.14.

By part(1) of Theorem 3.2.1, I have $\sup_{\tau\in[0,1]}|\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)| = O_P(h^2+\sqrt{\log T/(Th)})$. Hence, by $Th^6 \to 0$, I have

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

In addition, by the proof of part (1) of Theorem 3.2.1, the uniform convergence results of Lemmas B.2.5–B.2.6 and the condition that $\boldsymbol{\beta}(\tau)$ has third-order derivative, I have

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

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

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

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

by Lemma B.2.1. The proof is now completed. □

Appendix C

Appendix for Chapter 4

I first give some technical tools in Appendix C.1, while the proofs of main results are provided in Appendix C.2. Some preliminary lemmas are collected in Appendix C.3, and the proofs of preliminary lemmas are collected in Appendix C.4. Appendix C.5 discusses several computational issues of the local linear ML estimation. In what follows, M and $O(1)$ always stand for some bounded constants, and may be different at each appearance.

C.1 Technical Tools

Projection Operator: Define the projection operator

$$\mathscr{P}_t(\cdot) = E[\cdot \mid \mathscr{F}_t] - E[\cdot \mid \mathscr{F}_{t-1}],$$

where $\mathscr{F}_t = \sigma(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$. By the Jensen's inequality and the stationarity of $\widetilde{\mathbf{x}}_t(\tau)$, for $l \geq 0$, I have

$$\begin{aligned}
\|\mathscr{P}_{t-l}(\widetilde{\mathbf{x}}_t(\tau))\|_r &= \|E[\widetilde{\mathbf{x}}_t(\tau) \mid \mathscr{F}_{t-l}] - E[\widetilde{\mathbf{x}}_t(\tau) \mid \mathscr{F}_{t-l-1}]\|_r \\
&= \|E[\widetilde{\mathbf{x}}_t(\tau) \mid \mathscr{F}_{t-l}] - E[\widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau) \mid \mathscr{F}_{t-l-1}]\|_r \\
&= \|E[\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau) \mid \mathscr{F}_{t-l}]\|_r \\
&\leq \|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau)\|_r = \delta_r^{\mathbf{x}(\tau)}(l),
\end{aligned}$$

where $\widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau)$ is a coupled version of $\widetilde{\mathbf{x}}_t(\tau)$ with ε_{t-l} replaced by ε_{t-l}^*.

The Class $\mathscr{H}(C, \boldsymbol{\chi}, M)$:

Recall that I have defined $\boldsymbol{\Theta}_r$ in Assumption 4.2.1. Let $\boldsymbol{\chi} = \{\chi_j\}_{j=1}^{\infty}$ be a sequence of non-negative real numbers with $|\boldsymbol{\chi}|_1 := \sum_{j=1}^{\infty} \chi_j < \infty$ and $M > 0$ be some finite constant. Let $|\mathbf{z}|_{\boldsymbol{\chi}} := \sum_{j=1}^{\infty} \chi_j |\mathbf{z}_j|$ for any $\mathbf{z} \in (\mathbb{R}^m)^{\infty}$ and $C \geq 1$, where $\mathbf{z}_j$ is the j^{th} column of $\mathbf{z}$. A function $g(\mathbf{z}, \boldsymbol{\vartheta}) : (\mathbb{R}^m)^{\infty} \times \boldsymbol{\Theta}_r \to \mathbb{R}$ is in class $\mathscr{H}(C, \boldsymbol{\chi}, M)$ if

$$\begin{aligned}
&\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_r} |g(\mathbf{0}, \boldsymbol{\vartheta})| \leq M, \\
&\sup_{\mathbf{z}} \sup_{\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}'} \frac{|g(\mathbf{z}, \boldsymbol{\vartheta}) - g(\mathbf{z}, \boldsymbol{\vartheta}')|}{|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'|(1 + |\mathbf{z}|_{\boldsymbol{\chi}}^C)} \leq M, \\
&\sup_{\boldsymbol{\vartheta}} \sup_{\mathbf{z} \neq \mathbf{z}'} \frac{|g(\mathbf{z}, \boldsymbol{\vartheta}) - g(\mathbf{z}', \boldsymbol{\vartheta})|}{|\mathbf{z} - \mathbf{z}'|_{\boldsymbol{\chi}}(1 + |\mathbf{z}|_{\boldsymbol{\chi}}^{C-1} + |\mathbf{z}'|_{\boldsymbol{\chi}}^{C-1})} \leq M.
\end{aligned}$$

If g is vector- or matrix-valued, $g \in \mathscr{H}(C, \boldsymbol{\chi}, M)$ means that every component of g is in

$\mathscr{H}(C, \boldsymbol{\chi}, M)$.

Analytical Gradient:

Let

$$\ell(\mathbf{x}, \mathbf{z}; \boldsymbol{\vartheta}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})) - \frac{1}{2}\log\det\left(\mathbf{M}(\mathbf{z}; \boldsymbol{\vartheta})\right),$$

where $\mathbf{M}(\mathbf{z}; \boldsymbol{\vartheta}) = \mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta})\mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta})^\top$. Then the first partial derivative is as follows:

$$\begin{aligned}\frac{\partial \ell(\mathbf{x}, \mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i} &= (\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i} \\ &\quad -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \frac{\partial \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i}(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})) \\ &\quad -\frac{1}{2}\mathrm{tr}\left(\mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})\frac{\partial \mathbf{M}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i}\right),\end{aligned} \tag{C.1.1}$$

where ϑ_i is the i^{th} element of $\boldsymbol{\vartheta}$.

By (C.1.1), the second partial derivative of $\ell(\mathbf{x}, \mathbf{z}; \boldsymbol{\vartheta})$ is given by

$$\begin{aligned}&\frac{\partial^2 \ell(\mathbf{x}, \mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \\ = \; &(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta}) \frac{\partial^2 \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \\ &-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \frac{\partial^2 \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})) \\ &+(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \left(\frac{\partial \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i}\frac{\partial \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_j} + \frac{\partial \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_j}\frac{\partial \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i}\right) \\ &-\left(\frac{\partial \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_j}\right)^\top \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})\frac{\partial \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i} \\ &-\frac{1}{2}\mathrm{tr}\left(\mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})\frac{\partial^2 \mathbf{M}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right) - \frac{1}{2}\mathrm{tr}\left(\frac{\partial \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_j}\frac{\partial \mathbf{M}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i}\right).\end{aligned} \tag{C.1.2}$$

C.2 Proofs of the Main Results

Proof of Proposition 4.2.1.

(1). To prove the first result, I consider an approximated p-Markov process defined by

$$\begin{aligned}\widetilde{\mathbf{x}}_{p,t}(\tau) &= \boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{p,t-1}(\tau), \ldots, \widetilde{\mathbf{x}}_{p,t-p}(\tau), \mathbf{0}, \ldots; \boldsymbol{\theta}(\tau)\right) \\ &\quad +\mathbf{H}\left(\widetilde{\mathbf{x}}_{p,t-1}(\tau), \ldots, \widetilde{\mathbf{x}}_{p,t-p}(\tau), \mathbf{0}, \ldots; \boldsymbol{\theta}(\tau)\right)\boldsymbol{\varepsilon}_t\end{aligned} \tag{C.2.1}$$

for $p \geq 1$, and

$$\widetilde{\mathbf{x}}_{0,t}(\tau) = \boldsymbol{\mu}\left(\mathbf{0}, \ldots; \boldsymbol{\theta}(\tau)\right) + \mathbf{H}\left(\mathbf{0}, \ldots; \boldsymbol{\theta}(\tau)\right)\boldsymbol{\varepsilon}_t.$$

Let $\mu_{p,r}(\tau) = \|\widetilde{\mathbf{x}}_{p,t}(\tau)\|_r$ and $\Delta_{p,r}(\tau) = \|\widetilde{\mathbf{x}}_{p+1,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\|_r$.

By construction, I immediately obtain

$$\begin{aligned}\mu_{p,r}(\tau) &\leq \|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{0,t}(\tau)\|_r + \mu_{0,r}(\tau) \\ &\leq \left(\sum_{j=1}^{p}\alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r\sum_{j=1}^{p}\beta_j(\boldsymbol{\theta}(\tau))\right)\mu_{p,r}(\tau) + \mu_{0,r}(\tau),\end{aligned}$$

where the second inequality follows from Assumption 4.2.1.

Recall that I have defined $\rho(\tau) := \sum_{j=1}^{\infty} \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \sum_{j=1}^{\infty} \beta_j(\boldsymbol{\theta}(\tau))$ in the body of this proposition. As $0 \leq \rho(\tau) < 1$ by Assumption 4.2.1, I have

$$\sup_{p\geq 0} \mu_{p,r}(\tau) \leq (1-\rho(\tau))^{-1}\mu_{0,r}(\tau) < \infty.$$

Similarly, I have

$$\begin{aligned}\Delta_{p,r}(\tau) \quad &\leq \quad \left(\sum_{j=1}^{p} \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \sum_{j=1}^{p} \beta_j(\boldsymbol{\theta}(\tau))\right) \Delta_{p,r}(\tau) \\ &\quad + \left(\alpha_{p+1}(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \beta_{p+1}(\boldsymbol{\theta}(\tau))\right) \|\widetilde{\mathbf{x}}_{p+1,t-p-1}(\tau)\|_r .\end{aligned}$$

Hence,

$$\Delta_{p,r}(\tau) \leq \left(\alpha_{p+1}(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \beta_{p+1}(\boldsymbol{\theta}(\tau))\right)(1-\rho(\tau))^{-2}\mu_{0,r}(\tau) \to 0$$

as $p \to \infty$.

According to the above development, I am readily to conclude that $\widetilde{\mathbf{x}}_{p,t}(\tau) \to \widetilde{\mathbf{x}}_t(\tau)$ as $p \to \infty$ in the space of $\mathbb{L}^r = \{\mathbf{x} \mid \|\mathbf{x}\|_r < \infty\}$. As a limit of strictly stationary process in $\mathbb{L}^r$, $\widetilde{\mathbf{x}}_t(\tau)$ is a stationary process and $\sup_{\tau\in[0,1]} \|\widetilde{\mathbf{x}}_t(\tau)\|_r < \infty$.

(2). Let $\{\boldsymbol{\varepsilon}_t^*\}$ be an independent copy of $\{\boldsymbol{\varepsilon}_t\}$. Similar to (C.2.1), I define the process $\{\widetilde{\mathbf{x}}_{p,t}^*(\tau)\}$, in which the difference is that I use $\boldsymbol{\varepsilon}_t$ when $t \neq 0$, and use $\boldsymbol{\varepsilon}_t^*$ when $t = 0$. In addition, define the process $\{\widetilde{\mathbf{x}}_t^*(\tau)\}$ as $\{\widetilde{\mathbf{x}}_t(\tau)\}$, in which again the difference is that I use $\boldsymbol{\varepsilon}_t$ when $t \neq 0$, and use $\boldsymbol{\varepsilon}_t^*$ when $t = 0$. Further define $u_t = \left\|\widetilde{\mathbf{x}}_{p,t}^*(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\right\|_r$.

By construction, $u_t = 0$ for $t < 0$, and $u_0 = \left\|\widetilde{\mathbf{x}}_{p,0}^*(\tau) - \widetilde{\mathbf{x}}_{p,0}(\tau)\right\|_r = O\left(\|\boldsymbol{\varepsilon}_0^* - \boldsymbol{\varepsilon}_0\|_r\right) = O(1)$. For $t > 0$, Assumption 4.2.1 gives that

$$\left\|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\right\|_r \leq \sum_{j=1}^{p} \left(\alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_j(\boldsymbol{\theta}(\tau))\right) \left\|\widetilde{\mathbf{x}}_{p,t-j}(\tau) - \widetilde{\mathbf{x}}_{p,t-j}^*(\tau)\right\|_r . \tag{C.2.2}$$

Since $\sum_{j=1}^{p} \left(\alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_j(\boldsymbol{\theta}(\tau))\right) \leq \rho(\tau) < 1$, by a recursion argument, I have $u_t \leq u_0$ for all t.

Now, let $v_t = \max_{k\geq t} u_k$. Using (C.2.2) and the fact that v_t is a nonincreasing sequence, I have $v_t \leq \rho(\tau) v_{t-p}$ for all $t \geq 1$. Then recursively $v_t \leq \rho(\tau)^{-\lfloor -t/p \rfloor} v_{t+p\lfloor -t/p \rfloor}$. Since $v_{t+p\lfloor -t/p \rfloor} \leq u_0$ and $-\lfloor -t/p \rfloor \geq t/p$, I have $u_t \leq v_t \leq \rho(\tau)^{t/p} u_0$, i.e., $\left\|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\right\|_r = O(\rho(\tau)^{t/p})$.

The proof of the first result gives

$$\|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\|_r \leq \sum_{j=p}^{\infty} \Delta_{p,r} \leq \frac{\mu_r(\tau)}{(1-\rho(\tau))^2} \sum_{j=p}^{\infty} \left(\alpha_{p+1}(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_{p+1}(\boldsymbol{\theta}(\tau))\right).$$

The same bound holds for the quantity $\|\widetilde{\mathbf{x}}_t^*(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\|_r$. Thus,

$$\begin{aligned}\|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_t^*(\tau)\|_r \quad &\leq \quad \|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\|_r + \|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\|_r + \|\widetilde{\mathbf{x}}_t^*(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\|_r \\ &= \quad O\left(\rho(\tau)^{t/p} + \sum_{j=p+1}^{\infty} \left(\alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_j(\boldsymbol{\theta}(\tau))\right)\right),\end{aligned}$$

which completes the proof. □

Proof of Proposition 4.2.2.

(1). Write

$$\begin{aligned}\|\widetilde{\mathbf{x}}_t(\tau)-\widetilde{\mathbf{x}}_t(\tau')\|_r &\leq \|\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\ldots;\boldsymbol{\theta}(\tau)\right)-\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau'),\ldots;\boldsymbol{\theta}(\tau')\right)\|_r \\ &\quad +\|\boldsymbol{\varepsilon}_t\|_r\|\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\ldots;\boldsymbol{\theta}(\tau)\right)-\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau'),\ldots;\boldsymbol{\theta}(\tau')\right)\|_r \\ &\leq \sum_{j=1}^{\infty}\left(\alpha_j(\tau)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau)\right)\|\widetilde{\mathbf{x}}_{t-j}(\tau)-\widetilde{\mathbf{x}}_{t-j}(\tau')\|_r \\ &\quad +M|\tau-\tau'|\sum_{j=1}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau')\|_r,\end{aligned}$$

where the second inequality follows from Assumption 4.2.1 and Assumption 4.2.2. In view of the stationarity of $\widetilde{\mathbf{x}}_t(\tau)$, rearranging the terms in the above inequality yields that

$$\|\widetilde{\mathbf{x}}_t(\tau)-\widetilde{\mathbf{x}}_t(\tau')\|_r\leq M(1-\rho(\tau))^{-1}|\tau-\tau'|\sum_{j=1}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau')\|_r=O(|\tau-\tau'|).$$

(2). Write

$$\begin{aligned}\|\mathbf{x}_t-\widetilde{\mathbf{x}}_t(\tau_t)\|_r &\leq \|\boldsymbol{\mu}\left(\mathbf{x}_{t-1},\mathbf{x}_{t-2},\ldots;\boldsymbol{\theta}(\tau_t)\right)-\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau_t),\widetilde{\mathbf{x}}_{t-2}(\tau_t),\ldots;\boldsymbol{\theta}(\tau_t)\right)\|_r \\ &\quad +\|\boldsymbol{\varepsilon}_t\|_r\|\mathbf{H}\left(\mathbf{x}_{t-1},\mathbf{x}_{t-2},\ldots;\boldsymbol{\theta}(\tau_t)\right)-\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau_t),\widetilde{\mathbf{x}}_{t-2}(\tau_t),\ldots;\boldsymbol{\theta}(\tau_t)\right)\|_r \\ &\leq \sum_{j=1}^{\infty}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_t)\right\|_r \\ &\leq \sum_{j=1}^{\infty}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j}\vee 0)\right\|_r \\ &\quad +\sum_{j=1}^{\infty}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j}\vee 0)-\widetilde{\mathbf{x}}_{t-j}(\tau_t)\right\|_r.\end{aligned}$$

As $\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j}\vee 0)\right\|_r=0$ for $j\geq t$, by the first result of this proposition, I have

$$\begin{aligned}\|\mathbf{x}_t-\widetilde{\mathbf{x}}_t(\tau_t)\|_r &\leq \sum_{j=1}^{t-1}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j})\right\|_r \\ &\quad +M\cdot T^{-1}\sum_{j=1}^{\infty}j\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right).\end{aligned}$$

In addition, as $\left\|\mathbf{x}_1-\widetilde{\mathbf{x}}_1(\tau_1)\right\|_r=O(T^{-1})$ and $\sup_{t\geq 2}\sum_{j=1}^{t-1}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)<1$, I have

$$\begin{aligned}\|\mathbf{x}_t-\widetilde{\mathbf{x}}_t(\tau_t)\|_r &\leq \sum_{j=1}^{t-1}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)O(T^{-1}) \\ &\quad +M\cdot T^{-1}\sum_{j=1}^{\infty}j\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)=O(T^{-1}).\end{aligned}$$

The proof is now complete. □

Proof of Theorem 4.2.1.

(1). First, I introduce a few notations to facilitate the development. Let $\widehat{\boldsymbol{\eta}}(\tau):=[\widehat{\boldsymbol{\theta}}(\tau)^\top,\widehat{\boldsymbol{\theta}}^\star(\tau)^\top]^\top$,

$\boldsymbol{\eta}(\tau) := [\boldsymbol{\theta}(\tau)^\top, h\boldsymbol{\theta}^{(1)}(\tau)^\top]^\top$, and $\mathscr{L}_\tau(\boldsymbol{\eta}) := \mathscr{L}_\tau(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ for $\boldsymbol{\eta} = [\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top]^\top$. Recall that I have defined $\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}$, and let $\boldsymbol{\nabla}_{\boldsymbol{\eta}}$ be defined similarly with respect to the elements of $\boldsymbol{\eta}$.

By the Taylor expansion, I have

$$\widehat{\boldsymbol{\eta}}(\tau) - \boldsymbol{\eta}(\tau) \;=\; -(\boldsymbol{\nabla}^2_{\boldsymbol{\eta}}\mathscr{L}_\tau(\overline{\boldsymbol{\eta}}))^{-1}\boldsymbol{\nabla}_{\boldsymbol{\eta}}\mathscr{L}_\tau(\boldsymbol{\eta}(\tau)),$$

with $\overline{\boldsymbol{\eta}}$ between $\widehat{\boldsymbol{\eta}}(\tau)$ and $\boldsymbol{\eta}(\tau)$. By Lemma C.3.3.4, I have

$$|\boldsymbol{\nabla}_{\boldsymbol{\eta}}\mathscr{L}_\tau(\boldsymbol{\eta}(\tau)) - \boldsymbol{\nabla}_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))| = O_P((Th)^{-1}),$$

where $\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) = T^{-1}\sum_{t=1}^T \ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau) + \boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau))K_h(\tau_t-\tau)$.

Then I consider $\boldsymbol{\nabla}_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))$. Since each element of $\boldsymbol{\theta}(\tau)$ is in $C^3[0,1]$, I have $\boldsymbol{\theta}(\tau_t) = \boldsymbol{\theta}(\tau) + \boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau) + \mathbf{r}(\tau_t)$, where $\mathbf{r}(\tau_t) = \frac{1}{2}\boldsymbol{\theta}^{(2)}(\tau)(\tau_t-\tau)^2 + \frac{1}{6}\boldsymbol{\theta}^{(3)}(\overline{\tau})(\tau_t-\tau)^3$ with $\overline{\tau}$ between τ_t and τ. Let $\widehat{\mathbf{K}}((\tau_t-\tau)/h) = [K((\tau_t-\tau)/h), (\tau_t-\tau)/hK((\tau_t-\tau)/h)]^\top$. By the Mean Value Theorem, I have

$$\begin{aligned}&\boldsymbol{\nabla}_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - \frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))\\ =\;& -\frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes\left[\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t) - u\mathbf{r}(\tau_t))\mathbf{r}(\tau_t)\right]\end{aligned}$$

with some $u\in[0,1]$. Since $\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell$ is in class $\mathscr{H}(3,\boldsymbol{\chi},M)$ by Lemma C.3.2, using Lemma C.3.8 and $|\tau_t-\tau|\le h$ yields

$$\|\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t) - u\mathbf{r}(\tau_t)) - \boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau), \widetilde{\boldsymbol{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\|_1 = O(h).$$

The above analyses plus Lemma C.3.5 reveal that

$$\begin{aligned}&\boldsymbol{\nabla}_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - \frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))\\ =\;& -\frac{1}{2}h^2\frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes\left[\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau), \widetilde{\boldsymbol{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\cdot\boldsymbol{\theta}^{(2)}(\tau)\left(\frac{\tau_t-\tau}{h}\right)^2\right] + O_P(h^3)\\ =\;& \frac{1}{2}h^2\int_{-\tau/h}^{(1-\tau)/h} K(u)[u^2, u^3]^\top \mathrm{d}u \otimes\left(-\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)\right) + O_P(h^3).\end{aligned}$$

By Lemmas C.3.4 and C.3.5, I have

$$\boldsymbol{\nabla}_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) \to_P 0 \quad\text{and}\quad \sup_{\boldsymbol{\eta}}|\boldsymbol{\nabla}^2_{\boldsymbol{\eta}}\mathscr{L}_\tau(\boldsymbol{\eta}) - E(\boldsymbol{\nabla}^2_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}))| \to_P 0.$$

Hence, I have $\boldsymbol{\nabla}^2_{\boldsymbol{\eta}}\mathscr{L}_\tau(\overline{\boldsymbol{\eta}}) \to_P \boldsymbol{\Sigma}(\tau)$ and thus for any $\tau\in[h,1-h]$, as $Th^7\to 0$, I have

$$\begin{aligned}&\sqrt{Th}\left(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) - \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\theta}^{(2)}(\tau)\right)\\ =\;& -\boldsymbol{\Sigma}^{-1}(\tau)\frac{1}{\sqrt{Th}}\sum_{t=1}^T K((\tau_t-\tau)/h)\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) + o_P(1).\end{aligned}$$

In addition, by Lemma C.3.1, I have $E\left(\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\right) = 0$. To prove this theorem, by the Cramer-Wold device, it suffices to show that for any unit vector $\mathbf{d}$,

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K((\tau_t-\tau)/h)\mathbf{d}^\top \nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t)) \to_D N\left(\mathbf{0},\widetilde{v}_0\mathbf{d}^\top\boldsymbol{\Omega}(\tau)\mathbf{d}\right).$$

Note that $\{\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))\}_t$ is a sequence of martingale differences, I prove the asymptotic normality by using the martingale central limit theorem. I first consider the convergence of conditional variance. Let $w_t(u)=\frac{1}{\sqrt{Th}}K((\tau_t-\tau)/h)\mathbf{d}^\top\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u),\widetilde{\boldsymbol{z}}_{t-1}(u);\boldsymbol{\theta}(u))$. By Lemma C.3.8, I have

$$\begin{aligned}
&\sum_{t=1}^{T}\|w_t(\tau_t)^2-w_t(\tau)^2\|_1\\
\leq\ & \frac{1}{Th}\sum_{t=1}^{T}K((\tau_t-\tau)/h)^2\|\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))-\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau),\widetilde{\boldsymbol{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))\|_2\\
&\times 2\sup_u\|\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u),\widetilde{\boldsymbol{z}}_{t-1}(u);\boldsymbol{\theta}(u))\|_2\\
=\ & O(h)=o(1).
\end{aligned}$$

In addition, by Proposition 4.2.1, $\{E\left[(\mathbf{d}^\top\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u),\widetilde{\boldsymbol{z}}_{t-1}(u);\boldsymbol{\theta}(u)))^2\mid\mathscr{F}_{t-1}\right]\}_{t=1}^T$ is a sequence of stationary variables and thus I have

$$\begin{aligned}
&\sum_{t=1}^{T}E\left(w_t(\tau)^2\mid\mathscr{F}_{t-1}\right)\\
=\ & \frac{1}{Th}\sum_{t=1}^{T}K((\tau_t-\tau)/h)^2E\left[(\mathbf{d}^\top\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u),\widetilde{\boldsymbol{z}}_{t-1}(u);\boldsymbol{\theta}(u)))^2\mid\mathscr{F}_{t-1}\right]\\
\to_P\ & \widetilde{v}_0\mathbf{d}^\top\boldsymbol{\Omega}(\tau)\mathbf{d}.
\end{aligned}$$

I next verify the Lindeberg condition. The sum $\sum_{t=1}^{T}E\left(w_t^2(\tau_t)I(|w_t(\tau_t)|>v)\right)$ is bounded by

$$ME\left(\sup_\tau|\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau),\widetilde{\boldsymbol{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))|^2 I(\sup_\tau|\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau),\widetilde{\boldsymbol{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))|>\sqrt{Th}v)\right),$$

which converges to zero since $\|\sup_\tau|\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau),\widetilde{\boldsymbol{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))|\|_2<\infty$ by Lemma C.3.8.3. The asymptotic normality is then obtained.

The proof of the first result is now complete.

(2). For notation simplicity, I abbreviate $\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}),\boldsymbol{\mu}(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}),\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})$ to $\ell,\boldsymbol{\mu},\mathbf{M}$ in what follows. Note that

$$\begin{aligned}
\mathrm{d}\ell &= (\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\mathrm{d}\boldsymbol{\mu}-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\mathrm{d}\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})-\frac{1}{2}\mathrm{tr}\{\mathbf{M}^{-1}\mathrm{d}\mathbf{M}\}\\
&= (\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta}+\frac{1}{2}((\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\otimes(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1})\frac{\partial\mathrm{vec}(\mathbf{M})}{\partial\boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta}\\
&\quad -\frac{1}{2}\mathrm{vec}(\mathbf{M}^{-1})^\top\frac{\partial\mathrm{vec}(\mathbf{M})}{\partial\boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta}.
\end{aligned}$$

Hence, I have

$$
\begin{aligned}
&\frac{\partial \ell}{\partial \boldsymbol{\vartheta}}\frac{\partial \ell}{\partial \boldsymbol{\vartheta}^\top}\\
=\;& \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top}+\frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}\mathrm{vec}(\mathbf{M}^{-1})\mathrm{vec}(\mathbf{M}^{-1})^\top\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&+\frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\otimes\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&+\frac{1}{2}\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\otimes(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&-\frac{1}{2}\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})\mathrm{vec}(\mathbf{M}^{-1})^\top\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&+\frac{1}{2}\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}[(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\otimes\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&-\frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})\otimes\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})]\mathrm{vec}(\mathbf{M}^{-1})^\top\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&-\frac{1}{2}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}\mathrm{vec}(\mathbf{M}^{-1})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top}\\
&-\frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}\mathrm{vec}(\mathbf{M}^{-1})[(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}\otimes(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}. \qquad \text{(C.2.3)}
\end{aligned}
$$

In addition, if $\boldsymbol{\varepsilon}_t$ is normal distributed, I have $E(\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t}^\top\otimes\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t}^\top)=2\mathbf{N}_m+\mathrm{vec}(\mathbf{I}_m)\mathrm{vec}(\mathbf{I}_m)^\top$ and $E(\mathbf{c}\boldsymbol{\varepsilon_t}^\top\otimes\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t}^\top)=\mathbf{0}$, where $\mathbf{c}$ is independent of $\boldsymbol{\varepsilon_t}$, $2\mathbf{N}_m=\mathbf{I}_{m^2}+\mathbf{K}_{mm}$ and $\mathbf{K}_{mm}$ is a commutation matrix. By (C.2.3), if $\boldsymbol{\varepsilon_t}$ is normal distributed, I have

$$
\begin{aligned}
&\boldsymbol{\Omega}(\tau)=E\left(\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))\cdot\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))^\top\right)\\
=\;& E\left(\frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))^\top}{\partial \boldsymbol{\vartheta}}\mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))\frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))}{\partial \boldsymbol{\vartheta}^\top}\right)\\
&+\frac{1}{2}E\left(\frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau)))^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))\right.\\
&\left.\otimes\mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))]\frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau)))}{\partial \boldsymbol{\vartheta}^\top}\right).
\end{aligned}
$$

Next, consider the Hessian matrix.

$$\begin{aligned}
\mathrm{d}^2\ell \quad = \quad & -\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}} \mathbf{M}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} - \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \otimes \mathbf{M}^{-1}) \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} \\
& + \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}})^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \otimes \mathbf{I}_m) \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} \\
& + \frac{1}{2} \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1} \otimes \mathbf{M}^{-1}] \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2} \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}})^\top}{\partial \boldsymbol{\vartheta}} (\mathrm{vec}(\mathbf{M}^{-1})^\top \otimes \mathbf{I}_d) \mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2} \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1} \otimes \mathbf{M}^{-1}] \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2} \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \otimes \mathbf{M}^{-1}) \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2} \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1} \otimes \mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})) \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2} \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1} \otimes \mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}] \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top} \mathrm{d}\boldsymbol{\vartheta} \\
& + \frac{1}{2} \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}})^\top}{\partial \boldsymbol{\vartheta}} (\mathrm{vec}(\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}) \otimes \mathbf{I}_d) \mathrm{d}\boldsymbol{\vartheta}. \qquad \text{(C.2.4)}
\end{aligned}$$

By (C.2.4), if $\boldsymbol{\varepsilon_t}$ is normal distributed, I have

$$\begin{aligned}
& \boldsymbol{\Sigma}(\tau) = E\left(\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}} \ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))\right) \\
= \quad & -E\left(\frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))^\top}{\partial \boldsymbol{\vartheta}} \mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau)) \frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))}{\partial \boldsymbol{\vartheta}^\top}\right) \\
& -\frac{1}{2} E\left(\frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau)))^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))\right. \\
& \left. \otimes \mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))] \frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau)))}{\partial \boldsymbol{\vartheta}^\top}\right).
\end{aligned}$$

Then I have $\boldsymbol{\Omega}(\tau) = -\boldsymbol{\Sigma}(\tau)$ if $\boldsymbol{\varepsilon_t}$ is normal distributed. The proof is now complete. □

Proof of Corollary 4.2.1.

By Lemma C.3.5 (2) and the proof of Theorem 4.2.1, I have

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

In addition, applying Lemma C.3.3 (4), Lemma C.3.5 (2) and Lemma C.3.3 (2) to $g = \boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}} \ell$, I have

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

as $h(\log T)^2 \to 0$ and $\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}} \ell \in \mathscr{H}(3, \boldsymbol{\chi}, M)$.

For $\widehat{\boldsymbol{\Omega}}(\tau)$, as $\boldsymbol{\nabla_\theta} \ell (\boldsymbol{\nabla_\theta} \ell)^\top \in \mathscr{H}(6, \boldsymbol{\chi}, M)$, here I use a different argument to prove the result, which leads to weaker moment conditions.

By Lemma C.3.3.4 and Lemma C.3.8.2, I have

$$\begin{aligned}&\widehat{\boldsymbol{\Omega}}(\tau)-(Th)^{-1}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau))\nabla_{\boldsymbol{\theta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau))^\top\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\\ =\;& O_P((Th)^{-1/2}h^{-1/2}(\log T)^{1/2}+(Th)^{-1})=o_P(1),\end{aligned}$$

where $\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)=K\left(\frac{\tau_t-\tau}{h}\right)/\left(T^{-1}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\right)$.

Define $g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\theta}(\tau)):=\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))^\top$. By Lemma C.3.4.1, I have $\sup_{\tau\in[0,1]}\delta_{q/2}^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(j)=O(j^{-(3/2+s)})$.

Define $\mathbf{S}_T(\tau)=\sum_{t=1}^{T}\left[g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau))-E(g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau)))\right]\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)$ and

$$\mathbf{S}_{k,T}=\sum_{t=1}^{k}\left[g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau))-E(g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau)))\right].$$

By partial summation, I have

$$\mathbf{S}_T(\tau)=\sum_{t=1}^{T-1}\left[\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)-\widehat{K}\left(\frac{\tau_{t+1}-\tau}{h}\right)\right]\mathbf{S}_{t,T}+\widehat{K}\left(\frac{1-\tau}{h}\right)\mathbf{S}_{T,T}.$$

Hence, I have $\sup_{\tau\in[0,1]}|\mathbf{S}_T(\tau)|\leq M\max_t|\mathbf{S}_{t,T}|$. Note that $\{\mathscr{P}_{t-j}g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})\}_t$ forms a sequence of martingale differences. By the Doob's $\mathbb{L}^q$ maximal inequality, Burkholder's inequality and the elementary inequality $(\sum_i|a_i|)^q\leq\sum_i|a_i|^q$ for $0<q\leq 1$, I obtain that

$$\begin{aligned}\|\max_t|\mathbf{S}_{t,T}|\|_{q/2} &\leq \sum_{l=0}^{\infty}\|\max_{t=1,\ldots,T}|\sum_{s=1}^{t}\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau))|\|_{q/2}\\ &\leq \sum_{l=0}^{\infty}\frac{q/2}{q/2-1}\|\sum_{s=1}^{T}\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau))\|_{q/2}\\ &\leq \sum_{l=0}^{\infty}\frac{q/2}{(q/2-1)^2}\left[E\left(\sum_{s=1}^{T}(\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau)))^2\right)^{q/4}\right]\\ &\leq \frac{q/2}{(q/2-1)^2}\sum_{l=0}^{\infty}\left(\sum_{s=1}^{T}\|\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau))\|_{q/2}^{q/2}\right)^{2/q}\\ &\leq \frac{q/2}{(q/2-1)^2}T^{2/q}\sum_{l=0}^{\infty}\sup_{\tau\in[0,1]}\delta_{q/2}^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(l)\end{aligned}$$

which shows that $\sup_{\tau\in[0,1]}|\frac{1}{Th}\mathbf{S}_T(\tau)|=O_P(T^{2/q-1}h^{-1})=o_P(1)$. The result then follows directly by Lemma C.3.3.2. □

Proof of Theorem 4.2.2.

I prove this theorem by applying Lemma C.3.9 to the weak Bahadur representation of $\widehat{\boldsymbol{\theta}}(\tau)$ given in Lemma C.3.7.

By Lemma C.3.7, I have

$$\begin{aligned}
\sup_{\tau\in[h,1-h]} & \Bigg|\mathbf{C}(\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau))-\frac{1}{2}h^2\widetilde{c}_2\mathbf{C}\boldsymbol{\theta}^{(2)}(\tau) \\
& -\frac{1}{T}\sum_{t=1}^{T}(-\mathbf{C}\boldsymbol{\Sigma}^{-1}(\tau))\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\Bigg| \\
= & \ O_P(\gamma_T+\beta_T h^2+h^3+(Th)^{-1})=o_P((Th\log T)^{-1/2}) \qquad \text{(C.2.5)}
\end{aligned}$$

as $Th^7\log T\to 0$ and $Th^2/(\log T)^4\to\infty$. In addition, by Lemmas C.3.8, C.3.4.2 and C.3.9, I have

$$\begin{aligned}
\lim_{T\to\infty}\Pr & \left(\sqrt{\frac{Th}{\widetilde{v}_0}}\sup_{\tau\in[h,1-h]}\left|\boldsymbol{\Sigma}_{\mathbf{C}}^{-1/2}(\tau)\frac{1}{T}\sum_{t=1}^{T}(-\mathbf{C}\boldsymbol{\Sigma}^{-1}(\tau))\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\right|\right. \\
& \left. -B(1/h)\le\frac{u}{\sqrt{2\log(1/h)}}\right)=\exp(-2\exp(-u)). \qquad \text{(C.2.6)}
\end{aligned}$$

By (C.2.5) and (C.2.6), the proof is complete. □

Proof of Corollary 4.2.2.

By the proof of Lemma C.3.7, I have

$$\begin{aligned}
\sup_{\tau\in[0,1]} & \Bigg|\widehat{\boldsymbol{\eta}}(\tau)-\boldsymbol{\eta}(\tau)-\frac{1}{2}h^2\begin{bmatrix}\widetilde{c}_{0,h}(\tau) & \widetilde{c}_{1,h}(\tau)\\ \widetilde{c}_{1,h}(\tau) & \widetilde{c}_{2,h}(\tau)\end{bmatrix}^{-1}\begin{bmatrix}\widetilde{c}_{2,h}(\tau)\\ \widetilde{c}_{3,h}(\tau)\end{bmatrix}\otimes\boldsymbol{\theta}^{(2)}(\tau) \\
& -\frac{1}{T}\sum_{t=1}^{T}K_h(\tau_t-\tau)\begin{bmatrix}\widetilde{c}_{0,h}(\tau) & \widetilde{c}_{1,h}(\tau)\\ \widetilde{c}_{1,h}(\tau) & \widetilde{c}_{2,h}(\tau)\end{bmatrix}^{-1}\begin{bmatrix}1\\ \frac{\tau_t-\tau}{h}\end{bmatrix}\otimes-\boldsymbol{\Sigma}^{-1}(\tau)\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))\Bigg| \\
= & \ O_P((Th)^{-1/2}h^{3/2}(\log T)^{1/2})+O(h^3).
\end{aligned}$$

Hence, I have

$$\begin{aligned}
\sup_{\tau\in[0,1]} & \left|\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau)-\frac{1}{2}h^2 b_h(\tau)\boldsymbol{\theta}^{(2)}(\tau)-\frac{1}{T}\sum_{t=1}^{T}-\boldsymbol{\Sigma}^{-1}(\tau)\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))\omega_{t,h}(\tau)\right| \\
= & \ O_P((Th)^{-1/2}h^{3/2}(\log T)^{1/2})+O(h^3).
\end{aligned}$$

By Lemma C.3.10, there exists i.i.d. k-dimensional standard normal variables $\mathbf{v}_1,...,\mathbf{v}_T$ such that

$$\begin{aligned}
\sup_{\tau\in[0,1]} & \left|\frac{1}{Th}\sum_{t=1}^{T}\omega_{t,h}(\tau)(\nabla_{\boldsymbol{\theta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))-\boldsymbol{\Omega}^{1/2}(\tau_t)\mathbf{v}_t)\right| \\
= & \ O_P\left(\frac{T^{\frac{q(s+3)-4}{2q(2s+3)-4}}(\log T)^{\frac{2(s+1)(q+1)}{q(2s+3)-2}}}{Th}\right)=O_P\left(\frac{(\log T)^2(hT^{\frac{qs+2}{q(2s+3)-2}})^{-1/2}}{(Th)^{1/2}(\log T)^{1/2}}\right) \\
= & \ O_P\left(\frac{(\log T)^2(hT^{\nu})^{-1/2}}{(Th\log T)^{1/2}}\right)
\end{aligned}$$

with $\nu=\frac{qs+2}{q(2s+3)-2}$. Since $\boldsymbol{\Omega}(\tau)$ is Lipschitz continuous and $\{\mathbf{v}_t\}_{t=1}^T$ is a sequence of i.i.d. normal variables, I have

$$\sup_{\tau\in[0,1]}\left|\frac{1}{Th}\sum_{t=1}^{T}\omega_{t,h}(\tau)(\mathbf{\Omega}^{1/2}(\tau)-\mathbf{\Omega}^{1/2}(\tau_t))\mathbf{v}_t\right|$$
$$= O_P\left(\frac{h(\log T)^{1/2}}{(Th)^{1/2}}\right)=O_P\left(\frac{h\log T}{(Th\log T)^{1/2}}\right).$$

Combining the above analyses, I then complete the proof.

□

Proof of Proposition 4.2.3.

Note that in this case $\ell, \nabla\ell, \nabla^2\ell$ is in class $\mathscr{H}(2,\chi,M)$ as $\mathbf{H}(\mathbf{z};\boldsymbol{\theta})\equiv\mathbf{H}(\mathbf{0};\boldsymbol{\theta})$ by Lemma C.3.2. Hence, I only need that the innovation process has $4+s$ moments for some $s>0$ compared to $6+s$ moments needed in Theorem 4.2.2.

Consider Assumptions 4.2.1–4.2.2 first. For notation simplicity, I ignore the time-varying intercept, and rewrite model (4.2.6) as

$$\widetilde{\mathbf{y}}_t(\tau)=\mathbf{\Gamma}(\tau)\widetilde{\mathbf{y}}_{t-1}(\tau)+\widetilde{\mathbf{u}}_t(\tau),$$

where $\widetilde{\mathbf{y}}_t(\tau)=[\widetilde{\boldsymbol{\eta}}_t^\top(\tau),...,\widetilde{\boldsymbol{\eta}}_{t-q+1}^\top(\tau),\widetilde{\mathbf{x}}_t^\top(\tau),...,\widetilde{\mathbf{x}}_{t-p+1}^\top(\tau)]^\top$,

$$\widetilde{\mathbf{u}}_t(\tau)=[\widetilde{\mathbf{x}}_t^\top(\tau),\mathbf{0}_{m(q-1)\times1}^\top,\widetilde{\mathbf{x}}_t^\top(\tau),\mathbf{0}_{m(p-1)\times1}^\top]^\top$$

and

$$\mathbf{\Psi}(\tau)=\begin{bmatrix}-\mathbf{B}_1(\tau)&\cdots&-\mathbf{B}_{q-1}(\tau)&-\mathbf{B}_q(\tau)&-\mathbf{A}_1(\tau)&\cdots&-\mathbf{A}_{p-1}(\tau)&-\mathbf{A}_p(\tau)\\ \mathbf{I}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m&\mathbf{0}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ \vdots&\ddots&\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\ \mathbf{0}_m&\cdots&\mathbf{I}_m&\mathbf{0}_m&\mathbf{0}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ &&&&\mathbf{0}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ &&\mathbf{0}&&\mathbf{I}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ &&&&\vdots&\ddots&\vdots&\vdots\\ &&&&\mathbf{0}_m&\cdots&\mathbf{I}_m&\mathbf{0}_m\end{bmatrix}.$$

Let $\mathbf{J}=[\mathbf{I}_m,\mathbf{0}_{m\times(m(p+q-1))}]$ and $\mathbf{H}=[\mathbf{I}_m,\mathbf{0}_{m(q-1)\times1}^\top,\mathbf{I}_m,\mathbf{0}_{m(p-1)\times1}^\top]^\top$, I have

$$\begin{aligned}\widetilde{\boldsymbol{\eta}}_t(\tau)&=\widetilde{\mathbf{x}}_t(\tau)+\sum_{j=1}^{\infty}(\mathbf{J}\mathbf{\Psi}^j(\tau)\mathbf{H})\widetilde{\mathbf{x}}_{t-j}(\tau)\\ \widetilde{\mathbf{x}}_t(\tau)&=\sum_{j=1}^{\infty}(-\mathbf{J}\mathbf{\Psi}^j(\tau)\mathbf{H})\widetilde{\mathbf{x}}_{t-j}(\tau)+\widetilde{\boldsymbol{\eta}}_t(\tau)\end{aligned}$$

and thus $\mathbf{\Gamma}_j(\tau)=-\mathbf{J}\mathbf{\Psi}^j(\tau)\mathbf{H}$. Then Assumption 4.2.1 is automatically met if $\sum_{j=1}^{\infty}|\mathbf{\Gamma}_j(\tau)|<1$. By using the property of block matrix determinants and $\det(\mathbf{B}_\tau(L))\neq0$ for all $|L|\leq1$ (this implies the maximum eigenvalue of left upper $mq\times mq$ matrix in $\mathbf{\Gamma}(\tau)$ is less than 1), it can be show that the maximum eigenvalue of $\mathbf{\Gamma}(\tau)$, denoted by ρ, is less than 1 uniformly over $\tau\in[0,1]$. Hence, I have $\alpha_j(\boldsymbol{\theta}(\tau))=|\mathbf{\Gamma}_j(\tau)|=O(\rho^j)$ and $\beta_j(\boldsymbol{\theta}(\tau))=0$. In addition, $|\mathbf{\Psi}^j-\mathbf{\Psi}'^j|=|\sum_{i=1}^{j-1}\mathbf{\Psi}^i(\mathbf{\Psi}-\mathbf{\Psi}')\mathbf{\Psi}'^{j-1-i}|=|\mathbf{\Psi}-\mathbf{\Psi}'|O(\rho^{j-1})$. Then Assumption 4.2.2 is met.

However, by using techniques which are more specific to the VARMA models, the condition $\sum_{j=1}^{\infty}|\mathbf{\Gamma}_j(\tau)|<1$ can be weaken to $\det(\mathbf{A}_\tau(L))\neq 0$ for all $|L|\leq 1$. Similar to the above analysis, I have $\widetilde{\mathbf{x}}_t(\tau)=\sum_{j=0}^{\infty}\mathbf{\Phi}_j(\tau)\widetilde{\boldsymbol{\eta}}_{t-j}(\tau)$ with $|\mathbf{\Phi}_j(\tau)|=O(\rho^j)$ as $\det(\mathbf{A}_\tau(L))\neq 0$ for all $|L|\leq 1$, which implies that $\|\widetilde{\mathbf{x}}_t(\tau)\|_r<\infty$ and $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k)=O(\rho^k)$.

For the identification conditions stated in Assumption 4.2.3, it is well known that the final form or echelon form is enough to ensure the uniqueness of the VARMA representation.

For verifying Assumption 4.2.4, one need the derivatives of $\mathbf{\Gamma}_j$. Define

$$\boldsymbol{\alpha}=-\mathrm{vec}(\mathbf{B}_1,...,\mathbf{B}_q,\mathbf{A}_1,...,\mathbf{A}_p).$$

Note that $\mathrm{dvec}(\mathbf{\Psi})=(\mathbf{I}_{m(p+q)}\otimes\mathbf{J}^\top)\mathrm{d}\boldsymbol{\alpha}$ and $\mathrm{dvec}(\mathbf{\Psi}^j)=(\mathbf{\Psi}^\top\otimes\mathbf{I}_{m(p+q)})\mathrm{dvec}(\mathbf{\Psi}^{j-1})+(\mathbf{I}_{m(p+q)}\otimes\mathbf{\Psi}^{j-1}\mathbf{J}^\top)\mathrm{d}\boldsymbol{\alpha}$, it is easy to show that

$$\frac{\partial\mathrm{vec}(\mathbf{\Gamma}_j)}{\partial\boldsymbol{\alpha}^\top}=-\sum_{i=0}^{j-1}\mathbf{H}^\top(\mathbf{\Psi}^\top)^{j-1-i}\otimes\mathbf{J}\mathbf{\Psi}^j(\tau)\mathbf{J}^\top.$$

Hence, I have $|\frac{\partial\mathrm{vec}(\mathbf{\Gamma}_j)}{\partial\boldsymbol{\alpha}^\top}|=O(\rho^{j-1})$ and $|\frac{\partial\mathrm{vec}(\mathbf{\Gamma}_j)}{\partial\boldsymbol{\alpha}^\top}-\frac{\partial\mathrm{vec}(\mathbf{\Gamma}_j')}{\partial\boldsymbol{\alpha}'^\top}|=|\mathbf{\Psi}-\mathbf{\Psi}'|O(\rho^{j-2})$. Similarly, I can verify the conditions imposed on second order derivatives.

The proof is now complete.

□

Proof of Proposition 4.2.4.

Since $\mathbf{H}(\mathbf{z};\boldsymbol{\theta})$ is a positive and diagonal matrix, I have

$$\begin{aligned}|\mathbf{H}(\mathbf{z};\boldsymbol{\theta})-\mathbf{H}(\mathbf{z}';\boldsymbol{\theta})| &= |(\mathbf{H}^2(\mathbf{z};\boldsymbol{\theta})-\mathbf{H}^2(\mathbf{z}';\boldsymbol{\theta}))\cdot(\mathbf{H}(\mathbf{z};\boldsymbol{\theta})+\mathbf{H}(\mathbf{z}';\boldsymbol{\theta}))^{-1}|\\ &\leq \sum_{j=1}^{\infty}|\mathbf{\Psi}_j(\tau)|^{1/2}\cdot|\mathbf{x}_j-\mathbf{x}_j'|.\end{aligned}$$

Then Assumption 4.2.1 is automatically met if $\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r\sum_{j=1}^{\infty}|\mathbf{\Psi}_j(\tau)|^{1/2}<1$. In addition, as $|\mathbf{\Psi}_j(\tau)|$ converges to zero with exponential rate and $\frac{\partial h_{i,t}^{1/2}}{\partial\theta_i}=\frac{1}{2}h_{i,t}^{-1/2}\frac{\partial h_{i,t}}{\partial\theta_i}$, similar to the proof of Proposition 4.2.3 I can easily verify Assumptions 4.2.2 and 4.2.4. For the identification conditions of the GARCH process, I refer readers to Proposition 3.4 of Jeantheau [1998], who proves that assuming the minimal representation is enough for ensuring Assumption 4.2.3 holds.

However, by using techniques which are more specific to the GARCH models, the condition $\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r\sum_{j=1}^{\infty}|\mathbf{\Psi}_j(\tau)|^{1/2}<1$ can be weaken to $\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r^2\sum_{j=1}^{\infty}|\mathbf{\Psi}_j(\tau)|<1$. Define $\widetilde{\mathbf{y}}_t(\tau)=\widetilde{\mathbf{x}}_t(\tau)\odot\widetilde{\mathbf{x}}_t(\tau)$ and $\widetilde{\mathbf{V}}_t(\tau)=\mathrm{diag}\left(\widetilde{\boldsymbol{\eta}}_t(\tau)\odot\widetilde{\boldsymbol{\eta}}_t(\tau)\right)$. I first prove the existence of $\|\widetilde{\mathbf{y}}_t(\tau)\|_{r/2}$ (which implies the existence of $\|\widetilde{\mathbf{x}}_t(\tau)\|_r$) as well as its weak dependence property by means of a chaotic expansion. Since $\widetilde{\mathbf{y}}_t(\tau)=\widetilde{\mathbf{V}}_t(\tau)\boldsymbol{\alpha}(\tau)+\sum_{j=1}^{\infty}\widetilde{\mathbf{V}}_t(\tau)\mathbf{\Psi}_j(\tau)\widetilde{\mathbf{y}}_{t-j}(\tau)$, by substitute $\widetilde{\mathbf{y}}_{t-j}(\tau)$ recursively, I have

$$\widetilde{\mathbf{y}}_t(\tau)=\widetilde{\mathbf{V}}_t(\tau)\left\{\boldsymbol{\alpha}(\tau)+\sum_{k=1}^{\infty}\sum_{j_1,..,j_k=1}^{\infty}\mathbf{\Psi}_{j_1}(\tau)\widetilde{\mathbf{V}}_{t-j_1}(\tau)\cdots\mathbf{\Psi}_{j_k}(\tau)\widetilde{\mathbf{V}}_{t-j_1-\cdots-j_k}(\tau)\boldsymbol{\alpha}(\tau)\right\}.$$

To prove the boundedness of $\|\widetilde{\mathbf{y}}_t(\tau)\|_{r/2}$, since $\{\widetilde{\mathbf{V}}_t(\tau)\}$ are independent random variables, it suffices to show that

$$\sum_{k=1}^{\infty}\sum_{j_1,..,j_k=1}^{\infty}\left\|\mathbf{\Psi}_{j_1}(\tau)\widetilde{\mathbf{V}}_{t-j_1}(\tau)\cdots\mathbf{\Psi}_{j_k}(\tau)\widetilde{\mathbf{V}}_{t-j_1-\cdots-j_k}(\tau)\boldsymbol{\alpha}(\tau)\right\|_{r/2}<\infty.$$

By using $\sup_{\tau\in[0,1]}|\boldsymbol{\alpha}(\tau)| < \infty$ and $\|\widetilde{\mathbf{V}}_t(\tau)\|_{r/2}\sum_{j=1}^{\infty}|\boldsymbol{\Psi}_j(\tau)| < 1$, I have

$$\begin{aligned}
&\sum_{k=1}^{\infty}\sum_{j_1,..,j_k=1}^{\infty}\left\|\boldsymbol{\Psi}_{j_1}(\tau)\widetilde{\mathbf{V}}_{t-j_1}(\tau)\cdots\boldsymbol{\Psi}_{j_k}(\tau)\widetilde{\mathbf{V}}_{t-j_1-\cdots-j_k}(\tau)\boldsymbol{\alpha}(\tau)\right\|_{r/2}\\
\le\ & \sup_{\tau\in[0,1]}|\boldsymbol{\alpha}(\tau)|\sum_{k=1}^{\infty}\sum_{j_1,..,j_k=1}^{\infty}|\boldsymbol{\Psi}_{j_1}(\tau)|\cdots|\boldsymbol{\Psi}_{j_k}(\tau)|\|\widetilde{\mathbf{V}}_t(\tau)\|_{r/2}^{k}\\
\le\ & \sup_{\tau\in[0,1]}|\boldsymbol{\alpha}(\tau)|\sum_{k=1}^{\infty}\left(\|\widetilde{\mathbf{V}}_t(\tau)\|_{r/2}\sum_{j=1}^{\infty}|\boldsymbol{\Psi}_j(\tau)|\right)^{k} < \infty.
\end{aligned}$$

Hence, I have $\|\widetilde{\boldsymbol{x}}_t(\tau)\|_r < \infty$. Next, I show that $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k) = O(\rho^k)$ for some $0<\rho<1$. Write

$$\widetilde{\mathbf{y}}_t(\tau) = \operatorname{diag}\left(\boldsymbol{\alpha}(\tau) + \sum_{j=1}^{\infty}\boldsymbol{\Psi}_j(\tau)\widetilde{\mathbf{y}}_{t-j}(\tau)\right)(\widetilde{\boldsymbol{\eta}}_t(\tau)\odot\widetilde{\boldsymbol{\eta}}_t(\tau)).$$

By using the same arguments as in the proof of Proposition 4.2.3, I have $\delta_r^{\widetilde{\mathbf{y}}(\tau)}(k) = O(\rho^k)$ since $\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r^2\sum_{j=1}^{\infty}|\boldsymbol{\Psi}_j(\tau)| < 1$ and $|\boldsymbol{\Psi}_j(\tau)| = O(\rho^j)$. Since $|a-b| \le |a^2-b^2|^{1/2}$ for $a\ge 0, b\ge 0$ and $\mathbf{H}(\cdot)$ is a positive diagonal matrix, for $t\ge 1$, I have

$$\begin{aligned}
\|\widetilde{\boldsymbol{x}}_t(\tau) - \widetilde{\boldsymbol{x}}_t^*(\tau)\|_r &\le \sum_{j=1}^{t}|\boldsymbol{\Psi}_j(\tau)|^{1/2}\cdot\|\widetilde{\mathbf{y}}_{t-j}(\tau) - \widetilde{\mathbf{y}}_{t-j}^*(\tau)\|_{r/2}^{1/2}\cdot\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r\\
&= \sum_{j=1}^{t}O(\rho^{j/2})O(\rho^{(t-j)/2}) = O(\rho'^t)
\end{aligned}$$

for some $0<\rho'<1$.

The proof is now complete.

□

C.3 Preliminary Lemmas

First, I define a few notations for better presentation. First, let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$, where $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are the same generic vectors as in (4.2.3). Let $\widehat{K}(\cdot)$ be a a kernel function being Lipschitz continuous and bounded on $[-1,1]$.

For $\tau\in[0,1]$ and $\boldsymbol{\eta}\in\mathbf{E}_T(r) = \boldsymbol{\Theta}_r\times(h\cdot\boldsymbol{\Theta}^{(1)})$, define

$$G_\tau(\boldsymbol{\eta}) := \frac{1}{Th}\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\left[g(\mathbf{y}_t, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h) - E(g(\mathbf{y}_t, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h))\right], \tag{C.3.1}$$

where $g(\cdot)\in\mathscr{H}(C,\boldsymbol{\chi},M)$ and $\mathbf{y}_t = (\mathbf{x}_t, \mathbf{z}_{t-1})$. Let $G_\tau^c(\boldsymbol{\eta})$, $\widetilde{G}_\tau(\boldsymbol{\eta})$ denote the same quantity but with $\mathbf{y}_t$ replaced by $\mathbf{y}_t^c = (\mathbf{x}_t, \mathbf{z}_{t-1}^c)$ or $\widetilde{\mathbf{y}}_t(\tau_t) = (\widetilde{\mathbf{x}}_t(\tau_t), \widetilde{\mathbf{z}}_{t-1}(\tau_t))$.

In addition, let

$$\begin{aligned}
\widetilde{B}_\tau(\boldsymbol{\eta}) &:= \frac{1}{Th}\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h),\\
\mathbf{m}_t^{(2)}(u,\tau) &:= \widehat{K}\left(\frac{\tau-u}{h}\right)g(\widetilde{\mathbf{y}}_t(u), \boldsymbol{\theta}(\tau) - v\mathbf{d}(u,\tau))\cdot\mathbf{d}(u,\tau)
\end{aligned}$$

where $\mathbf{d}(u,\tau) := \boldsymbol{\theta}(u) - \boldsymbol{\theta}(\tau) - (u-\tau)\boldsymbol{\theta}^{(1)}(\tau)$ and some $v \in [0,1]$.

Lemma C.3.1. Suppose Assumptions 4.2.1 and 4.2.3 hold. Then, $E\left(\ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau); \boldsymbol{\vartheta})\right)$ is uniquely maximized at $\boldsymbol{\theta}(\tau)$.

Lemma C.3.2. Suppose Assumptions 4.2.3–4.2.4 hold. Then, $\ell, \nabla\ell, \nabla^2\ell \in \mathscr{H}(3, \boldsymbol{\chi}, M)$ for some $M > 0$ and $\boldsymbol{\chi} = \{\chi_j\}_{j=1,2,\ldots}$ with $\chi_j = O(j^{-(2+s)})$ and $s > 0$. In addition, if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) \equiv \mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$, $\ell, \nabla\ell, \nabla^2\ell \in \mathscr{H}(2, \boldsymbol{\chi}, M)$.

Lemma C.3.3. Suppose Assumptions 4.2.1–4.2.2 hold with $r \geq C$. Then

1. $\sup_{\tau\in[0,1]} \left\| \sup_{\boldsymbol{\eta}\neq\boldsymbol{\eta}'} \frac{|\widetilde{G}_\tau(\boldsymbol{\eta}) - \widetilde{G}_\tau(\boldsymbol{\eta}')|}{|\boldsymbol{\eta}-\boldsymbol{\eta}'|} \right\|_1 \leq M$ and $\left\| \sup_{\tau\neq\tau', \boldsymbol{\eta}\neq\boldsymbol{\eta}'} \frac{|\widetilde{G}_\tau(\boldsymbol{\eta}) - \widetilde{G}_{\tau'}(\boldsymbol{\eta}')|}{|\tau-\tau'|+|\boldsymbol{\eta}-\boldsymbol{\eta}'|} \right\|_1 \leq Mh^{-2}$;
2. $\sup_{\tau\in[0,1], \boldsymbol{\eta}\in\mathbf{E}_T(r)} |E(\widetilde{B}_\tau(\boldsymbol{\eta})) - \int_{-\tau/h}^{(1-\tau)/h} \widehat{K}(u) E(g(\widetilde{\mathbf{y}}_0(\tau), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u)) \mathrm{d}u| = O((Th)^{-1} + h)$;
3. $\left\| \sup_{\tau\neq\tau'} \frac{|\boldsymbol{\Pi}(\tau) - \boldsymbol{\Pi}(\tau')|}{|\tau-\tau'|} \right\|_1 \leq Mh^{-2}$ with $\boldsymbol{\Pi}(\tau) = (Th)^{-1} \sum_{t=1}^T \left[\mathbf{m}_t^{(2)}(\tau,\tau_t) - E(\mathbf{m}_t^{(2)}(\tau,\tau_t))\right]$.

In addition, suppose $\chi_j = O(j^{-(2+s)})$ for some $s > 0$, then

4. $\| \sup_{\tau\in[0,1], \boldsymbol{\eta}\in\mathbf{E}_T(r)} |\widetilde{G}_\tau(\boldsymbol{\eta}) - G_\tau^c(\boldsymbol{\eta})| \|_1 = O((Th)^{-1})$.

Lemma C.3.4. Let $g(\cdot) \in \mathscr{H}(C, \boldsymbol{\chi}, M)$, where $\chi_j = O(j^{-(a+s)})$ for some $s > 0$ and $a \geq 1$. Suppose Assumptions 4.2.1–4.2.2 hold with $q = r/C \geq 1$, and

$$\sup_{\tau\in[0,1]} [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(a+1+s)}).$$

Then I obtain

1. $\sup_{\tau\in[0,1]} \delta_q^{\sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau), \boldsymbol{\vartheta})|}(j) = O(j^{-(a+s)})$;
2. $\sup_{\tau\in[0,1]} \sup_{u,\boldsymbol{\eta}} \delta_q^{m(\tau,\boldsymbol{\eta},u)}(j) = O(j^{-(a+s)})$ and $\sup_{\tau\in[0,1]} \delta_q^{\sup_{u,\boldsymbol{\eta}} |m(\tau,\boldsymbol{\eta},u)|}(j) = O(j^{-(a+s)})$, where $m_t(\tau,\boldsymbol{\eta},u) := \widehat{K}\left(\frac{\tau-u}{h}\right) g(\widetilde{\mathbf{y}}_t(\tau), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau-u)/h)$;
3. $\sup_{\tau,u\in[0,1]} \delta_q^{\mathbf{m}_t^{(2)}(u,\tau)}(j) = O(h^2 j^{-(a+s)})$ and $\sup_{u\in[0,1]} \delta_q^{\sup_\tau |\mathbf{m}_t^{(2)}(u,\tau)|}(j) = O(h^2 j^{-(a+s)})$.

Lemma C.3.5. Under the conditions of Lemma C.3.4 with $q = r/C > 1$, then

1. $\|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q = O\left((Th)^{-(q'-1)/q'}\right)$ with $q' = \min(2,q)$,
2. $\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)} |\widetilde{G}_\tau(\boldsymbol{\eta})| = o_P(1)$;

Suppose further $q = r/C > 2$ and $a \geq 3/2$. Then

3. $\sup_{\tau\in[0,1]} \sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)} |\widetilde{G}_\tau(\boldsymbol{\eta})| = O_P((\log T)^{1/2}(Th)^{-1/2}h^{-1/2})$.

Lemma C.3.6. Suppose Assumptions 4.2.1–4.2.5 hold with $r > 6$, and

$$\sup_{\tau\in[0,1]} [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(a+1+s)})$$

for $a \geq 3/2$ and some $s > 0$. Then

$$\sup_{\tau\in[0,1]} \Big| \nabla\widetilde{\mathscr{L}}_\tau(\boldsymbol{\theta}(\tau), h\boldsymbol{\theta}^{(1)}(\tau)) - E[\nabla\widetilde{\mathscr{L}}_\tau(\boldsymbol{\theta}(\tau), h\boldsymbol{\theta}^{(1)}(\tau))]$$
$$- \frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t-\tau)/h) \otimes \nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \Big| = O_P(h^2\beta_T),$$

where

$$\begin{aligned}\beta_T &= (\log T)^{1/2}(Th)^{-1/2}h^{-1/2},\\ \widehat{\mathbf{K}}((\tau_t-\tau)/h) &= K((\tau_t-\tau)/h)[1,(\tau_t-\tau)/h]^\top,\\ \widetilde{\mathscr{L}}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau)) &:= T^{-1}\sum_{t=1}^{T}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau)+\boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau))K_h(\tau_t-\tau).\end{aligned}$$

Lemma C.3.7. Under the conditions of Theorem 4.2.2,

$$(1).\qquad \sup_{\tau\in[h,1-h]}\left|-\boldsymbol{\Sigma}(\tau)(\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau))-\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\mathscr{L}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau))\right|=O_P(\gamma_T),$$

$$(2).\qquad \sup_{\tau\in[h,1-h]}\left|\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\mathscr{L}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau))+\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)\right.$$
$$\left.-\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\right|=O_P(\beta_Th^2+h^3+(Th)^{-1}),$$

where $\beta_T=(Th)^{-1/2}h^{-1/2}(\log T)^{1/2}$ and $\gamma_T=(\beta_T+h)((Th)^{-1/2}\log T+h^2)$.

C.3.1 Secondary Lemmas

Before proceeding further, I introduce some extra notations. Assume that there exists some measurable function $\widetilde{\mathbf{H}}(\cdot,\cdot)$ such that for $\forall\tau\in[0,1]$, $\widetilde{\mathbf{h}}_t(\tau)=\widetilde{\mathbf{H}}(\tau,\mathscr{F}_t)\in\mathbb{R}^d$ is well defined, where $\mathscr{F}_t=\sigma(\boldsymbol{\varepsilon}_t,\boldsymbol{\varepsilon}_{t-1},\ldots)$. Let

$$\widetilde{\mathbf{D}}_{\widetilde{\mathbf{h}}}(\tau):=(Th)^{-1}\sum_{t=1}^{T}\widetilde{\mathbf{h}}_t(\tau_t)\widehat{K}((\tau_t-\tau)/h)\quad\text{and}\quad \boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}(\tau)=\sum_{j=-\infty}^{\infty}E[\widetilde{\mathbf{h}}_0(\tau)\widetilde{\mathbf{h}}_j^\top(\tau)].$$

Assume that $\boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}(\tau)$ is Lipschitz continuous and its smallest eigenvalue is bounded away from 0 uniformly over $\tau\in[0,1]$. In what follows, I let $\widetilde{h}_{0,i}(\tau)$ stand for the i^{th} component of $\widetilde{\mathbf{h}}_t(\tau)$.

Lemma C.3.8. Let $q>0$. Let $g\in\mathscr{H}(C,\boldsymbol{\chi},M)$. Let $\mathbf{y}=(\mathbf{y}_0,\mathbf{y}_1,\mathbf{y}_2,\ldots)$ and $\mathbf{y}'=(\mathbf{y}'_0,\mathbf{y}'_1,\mathbf{y}'_2,\ldots)$ be two sequences of random variables. Assume that $\max_{j\geq 0}\|\mathbf{y}_j\|_{qC}\leq M$ and $\max_{j\geq0}\|\mathbf{y}'_j\|_{qC}\leq M$. Then, I have

1. $\left\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r}|g(\mathbf{y},\boldsymbol{\vartheta})-g(\mathbf{y}',\boldsymbol{\vartheta})|\right\|_q\leq M\sum_{j=0}^{\infty}\chi_j\|\mathbf{y}_j-\mathbf{y}'_j\|_{qC}$;
2. $\left\|\sup_{\boldsymbol{\vartheta}\neq\boldsymbol{\vartheta}'}\frac{|g(\mathbf{y},\boldsymbol{\vartheta})-g(\mathbf{y},\boldsymbol{\vartheta}')|}{|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|}\right\|_q\leq M$;
3. $\left\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r}|g(\mathbf{y},\boldsymbol{\vartheta})|\right\|_q\leq M$.

Lemma C.3.8 is Lemma D.4 of Karmakar et al. [forthcoming].

Lemma C.3.9. Assume that for $i=1,\ldots,d$

1. $\sup_{\tau\in[0,1]}\|\widetilde{h}_{0,i}(\tau)\|_q<\infty$ with some $2\leq q\leq 4$,
2. $\sup_{\tau\neq\tau'}\|\widetilde{h}_{0,i}(\tau)-\widetilde{h}_{0,i}(\tau')\|_2/|\tau-\tau'|<\infty$,
3. $\sup_{\tau\in[0,1]}\delta_q^{\widetilde{h}_{0,i}(\tau)}(j)=O(j^{-(2+s)})$ for some $s\geq 0$.

In addition, assume that $h(\log T)^{3/2} \to 0$ and $\frac{(\log T)^4}{T^{(sq+2)/(2sq+3q-2)}h} \to 0$. Then

$$\lim_{T\to\infty} \Pr\left(\sqrt{\frac{Th}{\widetilde{v}_0}} \sup_{\tau\in[h,1-h]} \left|\boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}^{-1/2}(\tau)\widetilde{\mathbf{D}}_{\widetilde{\mathbf{h}}}(\tau)\right| - B(m^*) \le \frac{u}{\sqrt{2\log(m^*)}}\right) = \exp(-2\exp(-u)),$$

where

$$\begin{aligned} B(m^*) &= \sqrt{2\log(m^*)} + \frac{\log(C_K) + (k/2 - 1/2)\log(\log(m^*)) - \log(2)}{\sqrt{2\log(m^*)}}, \\ C_K &= \frac{\{\int_{-1}^{1} |K^{(1)}(u)|^2 \mathrm{d}u/\widetilde{v}_0\pi\}^{1/2}}{\Gamma(k/2)}, \quad m^* = 1/h, \end{aligned}$$

and $\Gamma(\cdot)$ is the Gamma function.

Lemma C.3.9 is Lemma B.3 of Karmakar et al. [forthcoming].

Lemma C.3.10. Assume that for $i = 1, \dots, d$

1. $\sup_{\tau\in[0,1]} \|\widetilde{h}_{0,i}(\tau)\|_q < \infty$ with some $2 \le q \le 4$,
2. $\sup_{\tau\ne\tau'} \|\widetilde{h}_{0,i}(\tau) - \widetilde{h}_{0,i}(\tau')\|_2/|\tau - \tau'| < \infty$,
3. $\sup_{\tau\in[0,1]} \delta_q^{\widetilde{h}_{0,i}(\tau)}(j) = O(j^{-(2+s)})$ for some $s \ge 0$.

Let $\mathbf{S}_{\widetilde{\mathbf{h}}}(t) = \sum_{s=1}^{t} \widetilde{\mathbf{h}}_s(\tau_s)$. Then on a richer probability space, there exists i.i.d. k-dimensional standard normal variables $\mathbf{v}_1, \mathbf{v}_2, \dots$ and a process $\mathbf{S}_{\widetilde{\mathbf{h}}}^0(t) = \sum_{s=1}^{t} \boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_s)\mathbf{v}_s$ such that

$$(\mathbf{S}_{\widetilde{\mathbf{h}}}(t))_{t=1}^{T} =_D (\mathbf{S}_{\widetilde{\mathbf{h}}}^0(t))_{t=1}^{T} \quad \text{and} \quad \max_{t\ge1} |\mathbf{S}_{\widetilde{\mathbf{h}}}(t) - \mathbf{S}_{\widetilde{\mathbf{h}}}^0(t)| = O_P(\pi_T)$$

where $\pi_T = T^{\frac{q(s+3)-4}{2q(2s+3)-4}}(\log T)^{\frac{2(s+1)(q+1)}{q(2s+3)-2}}$.

Lemma C.3.10 is from Theorem 1 and Corollary 2 of Wu and Zhou [2011].

C.4 Proofs of Preliminary Lemmas

Proof of Lemma C.3.1.

Let

$$\begin{aligned} \mathbf{M}_t(\boldsymbol{\vartheta}, \boldsymbol{\theta}(\tau)) \;:=\; & (\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta})^\top)^{-1/2}\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))^\top \\ & \times(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta})^\top)^{-1/2}. \end{aligned}$$

By Assumption 4.2.1.1 and the construction of $\widetilde{\mathbf{x}}_t(\tau)$, I write

$$\begin{aligned}
& E\left(\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\right) \\
= & -\frac{1}{2}E\log\det\left\{\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top\right\} \\
& -\frac{1}{2}E\mathrm{tr}\left\{(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}[\widetilde{\mathbf{x}}_t(\tau)-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})][\widetilde{\mathbf{x}}_t(\tau)-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top\right\} \\
= & -\frac{1}{2}E\log\det\left\{\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top\right\}-\frac{1}{2}E\mathrm{tr}\left\{\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau))\right\} \\
& -\frac{1}{2}E\left([\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}\right. \\
& \left.\times[\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]\right) \\
= & -\frac{1}{2}\left[-E\log\det\left(\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau))\right)+E\mathrm{tr}\left\{\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau))\right\}\right] \\
& -\frac{1}{2}E\log\det\left(\widetilde{\mathbf{H}}_t(\tau,\boldsymbol{\theta}(\tau))\widetilde{\mathbf{H}}_t(\tau,\boldsymbol{\theta}(\tau))^\top\right) \\
& -\frac{1}{2}E\left([\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}\right. \\
& \left.\times[\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]\right).
\end{aligned}$$

For any positive definite matrix $\mathbf{M}$ with eigenvalues $\lambda_1,\ldots,\lambda_m>0$, I have

$$f(\mathbf{M}):=-\log\det\left(\mathbf{M}\right)+\mathrm{tr}\left\{\mathbf{M}\right\}=\sum_{i=1}^m(\lambda_i-\log\lambda_i)\geq m,$$

where the equality holds if $\lambda_1=\cdots=\lambda_m=1$ in which case $\mathbf{M}=\mathbf{I}_m$. Thus, $f(\mathbf{M})$ is uniquely minimized at $\mathbf{M}=\mathbf{I}_m$, which implies that $E[f(\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau)))]$ is uniquely minimized at $\boldsymbol{\vartheta}=\boldsymbol{\theta}(\tau)$ by Assumption 4.2.3.2. In addition, since $\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top$ is a positive definite matrix, then

$$\begin{aligned}
& E\left([\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}\right. \\
& \left.\times[\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]\right)\geq 0
\end{aligned}$$

is uniquely minimized at $\boldsymbol{\vartheta}=\boldsymbol{\theta}(\tau)$ by Assumption 4.2.3.2. Hence, $E\left(\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\right)$ is uniquely maximized at $\boldsymbol{\theta}(\tau)$. □

Proof of Lemma C.3.2.

I first consider $\ell(\cdot)$. Write

$$\begin{aligned}
& \ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta})-\ell(\mathbf{x}',\mathbf{z}';\boldsymbol{\vartheta}) \\
= & -\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))-(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))^\top\mathbf{M}^{-1}(\mathbf{z}';\boldsymbol{\vartheta})(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))\right] \\
& -\frac{1}{2}\left[\log\det\left(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})\right)-\log\det\left(\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta})\right)\right] \\
:= & -\frac{1}{2}(I_1+I_2),
\end{aligned}$$

where the definitions of I_1 and I_2 should be obvious, and $\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})=\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})^\top$.

For I_2, I have

$$\begin{aligned}
|\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta})| & \leq |\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})|\left(|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})|+|\mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})|\right) \\
& \leq M|\mathbf{z}-\mathbf{z}'|_{\boldsymbol{\chi}}(2+|\mathbf{z}|_{\boldsymbol{\chi}}+|\mathbf{z}'|_{\boldsymbol{\chi}})
\end{aligned}$$

where the second inequality follows from the facts that

$$|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) - \mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})| = O(|\mathbf{z}-\mathbf{z}'|_{\boldsymbol{\chi}}),$$
$$|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})| \le |\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) - \mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})| + |\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})| = O(1+|\mathbf{z}|_{\boldsymbol{\chi}})$$

by using Assumption 4.2.1.2 twice.

By Assumption 4.2.3, it is easy to know that $\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})) \ge \underline{H} > 0$, which in connection with the fact $\log(\cdot)$ is Lipschitz continuous on $[\underline{H},\infty)$ yields that

$$I_2 \le M|\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})) - \det\left(\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta})\right)|.$$

In addition, for an invertible matrix $\mathbf{A}$, $\det(\mathbf{A}+\mathbf{B}) = \det(\mathbf{A}) + \mathrm{tr}(\mathbf{A}^{-1,\top}\mathbf{B}) + o(|\mathbf{B}|)$, and for a positive definite matrix $\mathbf{A}$ and symmetric matrix $\mathbf{B}$, $|\mathrm{tr}(\mathbf{A}^{-1,\top}\mathbf{B})| \le |\mathbf{B}|\mathrm{tr}(\mathbf{A}^{-1})$. Hence, I have

$$\begin{aligned}
I_2 &\le M|\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})) - \det\left(\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta})\right)| \\
&\le M\mathrm{tr}(\mathbf{M}^{-1}(\mathbf{z}';\boldsymbol{\vartheta}))\cdot|\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}) - \mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta})| \\
&= O(|\mathbf{z}-\mathbf{z}'|_{\boldsymbol{\chi}}(1+|\mathbf{z}|_{\boldsymbol{\chi}}+|\mathbf{z}'|_{\boldsymbol{\chi}})).
\end{aligned}$$

Note that if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) \equiv \mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$, $I_2 = 0$.

For I_1, since $|\mathbf{H}^{-1}(\mathbf{z};\boldsymbol{\vartheta})|$ is bounded by Assumption 4.2.3, I can obtain that

$$\begin{aligned}
I_1 &\le |\mathbf{H}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})) - \mathbf{H}^{-1}(\mathbf{z}';\boldsymbol{\vartheta})(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))| \\
&\quad \cdot(|\mathbf{H}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))| + |\mathbf{H}^{-1}(\mathbf{z}';\boldsymbol{\vartheta})(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))|) \\
&= O(|\mathbf{y}-\mathbf{y}'|_{\boldsymbol{\chi}}\cdot(1+|\mathbf{y}|_{\boldsymbol{\chi}}^2+|\mathbf{y}'|_{\boldsymbol{\chi}}^2)),
\end{aligned}$$

where $\mathbf{y} = (\mathbf{x},\mathbf{z})$. Similarly, if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) \equiv \mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$, $I_1 = O(|\mathbf{y}-\mathbf{y}'|_{\boldsymbol{\chi}}\cdot(1+|\mathbf{y}|_{\boldsymbol{\chi}}+|\mathbf{y}'|_{\boldsymbol{\chi}}))$.

For $\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}) - \ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}')$, write

$$\begin{aligned}
&\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}) - \ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}') \\
= & -\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})) - (\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}'))^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta}')(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}'))\right] \\
& -\frac{1}{2}\left[\log\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})) - \log\det\left(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}')\right)\right] \\
:= & -\frac{1}{2}(I_3+I_4).
\end{aligned}$$

Similar to the development for I_1 and I_2, I can obtain that

$$I_3 = O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|(1+|\mathbf{y}|_{\boldsymbol{\chi}}^3) \quad \text{and} \quad I_4 = O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|(1+|\mathbf{z}|_{\boldsymbol{\chi}}^2)),$$

where I again let $\mathbf{y} = (\mathbf{x},\mathbf{z})$. Also if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) \equiv \mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$,

$$I_3 = O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|(1+|\mathbf{y}|_{\boldsymbol{\chi}}^2) \quad \text{and} \quad I_4 = O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|).$$

Combing the above analysis, I have shown $\ell \in \mathscr{H}(3,\boldsymbol{\chi},M)$. In addition, if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) \equiv \mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$, $\ell \in \mathscr{H}(2,\boldsymbol{\chi},M)$.

Similar to the development for ℓ, I can show $\boldsymbol{\nabla}\ell$, $\boldsymbol{\nabla}^2\ell \in \mathscr{H}(3,\boldsymbol{\chi},M)$ and $\boldsymbol{\nabla}\ell, \boldsymbol{\nabla}^2\ell \in \mathscr{H}(2,\boldsymbol{\chi},M)$ if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta}) \equiv \mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$.

The proof is now complete. □

Proof of Lemma C.3.3.

(1). By Proposition 4.2.1.1, I have $\sup_{\tau\in[0,1]}\|\widetilde{\mathbf{x}}_t(\tau)\|_C<\infty$. Since $g\in\mathscr{H}(C,\boldsymbol{\chi},M)$, I have

$$\sup_{\boldsymbol{\eta}\neq\boldsymbol{\eta}'}\frac{|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|}{|\boldsymbol{\eta}-\boldsymbol{\eta}'|}\leq M(Th)^{-1}\sum_{t=1}^T\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\left[2+|\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C+\||\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C\|_1\right].$$

Using $(Th)^{-1}\sum_{t=1}^T\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)<\infty$, I have

$$\left\|\sup_{\boldsymbol{\eta}\neq\boldsymbol{\eta}'}\frac{|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|}{|\boldsymbol{\eta}-\boldsymbol{\eta}'|}\right\|_1\leq M\max_t\||\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C\|_1<\infty.$$

In addition, by using the Lipschitz property of $\widehat{K}(\cdot)$, I have

$$\begin{aligned}
&|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_{\tau'}(\boldsymbol{\eta}')|\\
\leq\ & (Th)^{-1}\sum_{t=1}^T\left|\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)-\widehat{K}\left(\frac{\tau_t-\tau'}{h}\right)\right|\cdot\sup_{\boldsymbol{\vartheta}}(|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})|+\|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})\|_1)\\
&+(Th)^{-1}\sum_{t=1}^T\widehat{K}\left(\frac{\tau_t-\tau'}{h}\right)\cdot|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2(\tau_t-\tau)/h)-g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1'+\boldsymbol{\eta}_2'(\tau_t-\tau')/h)|\\
\leq\ & M\left(h^{-2}|\tau-\tau'|+h^{-1}|\boldsymbol{\eta}-\boldsymbol{\eta}'|+h^{-2}|\boldsymbol{\eta}_2|\cdot|\tau-\tau'|\right)\cdot\frac{1}{T}\sum_{t=1}^T(2+|\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C+\||\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C\|_1).
\end{aligned}$$

Combing the above analyses, the first result follows.

(2). By Lemma C.3.8.1 and Proposition 4.2.2, for $|\tau_t-\tau|\leq h$, I have

$$\begin{aligned}
&\|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)-g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\|_1\\
\leq\ & M\sum_{j=0}^\infty\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau_t)-\widetilde{\mathbf{x}}_{t-j}(\tau)\|_C=O(h).
\end{aligned}$$

Hence, I have

$$\begin{aligned}
&\left\|\widetilde{B}_\tau(\boldsymbol{\eta})-\frac{1}{Th}\sum_{t=1}^T\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\right\|_1\\
\leq\ & M\frac{1}{Th}\sum_{t=1}^T\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\sum_{j=0}^\infty\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau_t)-\widetilde{\mathbf{x}}_{t-j}(\tau)\|_C=O(h)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{Th}\sum_{t=1}^T\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)E\left[g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\right]\\
=\ & \int_{-\tau/h}^{(1-\tau)/h}\widehat{K}(u)E(g(\widetilde{\mathbf{y}}_0(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2u))\mathrm{d}u+O((Th)^{-1})
\end{aligned}$$

by the definition of Riemann integral and the stationarity of $\widetilde{\mathbf{y}}_t(\tau)$.

(3). Write

$$
\begin{aligned}
& |\mathbf{m}_t^{(2)}(u,\tau) - \mathbf{m}_t^{(2)}(u,\tau')| \\
\leq\ & |\widehat{K}\left(\frac{\tau-u}{h}\right) - \widehat{K}\left(\frac{\tau'-u}{h}\right)| \cdot |g(\widetilde{\mathbf{y}}_t(u), \boldsymbol{\theta}(\tau) - v\mathbf{d}(u,\tau))| \cdot |\mathbf{d}(u,\tau)| \\
& + |\widehat{K}\left(\frac{\tau'-u}{h}\right)| \cdot |g(\widetilde{\mathbf{y}}_t(u), \boldsymbol{\theta}(\tau) - v\mathbf{d}(u,\tau)) - g(\widetilde{\mathbf{y}}_t(u), \boldsymbol{\theta}(\tau') - v\mathbf{d}(u,\tau'))| \cdot |\mathbf{d}(u,\tau)| \\
& + |\widehat{K}\left(\frac{\tau'-u}{h}\right)| \cdot |g(\widetilde{\mathbf{y}}_t(u), \boldsymbol{\theta}(\tau') - v\mathbf{d}(u,\tau'))| \cdot |\mathbf{d}(u,\tau) - \mathbf{d}(u,\tau')| \\
:=\ & I_1 + I_2 + I_3.
\end{aligned}
$$

By the Lipschitz continuity of $\widehat{K}(\cdot)$ and $\|\sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(u), \boldsymbol{\vartheta})|\|_1 = O(1)$ (by Lemma C.3.8.3), I have

$$E(I_1) = O(h^{-1}|\tau - \tau'|).$$

Similarly, by Lemma C.3.8.2 and $|\mathbf{d}(u,\tau)| = O(1)$, I have

$$E(I_2) = O(|\tau - \tau'|).$$

By the Lipschitz continuity of $\mathbf{d}(u,\cdot)$, I have $E(I_3) = O(|\tau - \tau'|)$. Hence,

$$\left|\frac{1}{Th}\sum_{t=1}^{T}[\mathbf{m}_t^{(2)}(\tau,\tau_t) - \mathbf{m}_t^{(2)}(\tau',\tau_t)]\right| \leq \frac{1}{Th}\sum_{t=1}^{T}|\mathbf{m}_t^{(2)}(\tau,\tau_t) - \mathbf{m}_t^{(2)}(\tau',\tau_t)| = O(h^{-2}|\tau-\tau'|).$$

The proof is now complete.

(4). By Propositions 4.2.1.1 and 4.2.2.2, I have $\sup_{\tau\in[0,1]} \|\widetilde{\mathbf{x}}_t(\tau)\|_C < \infty$ and $\max_t \|\mathbf{x}_t - \widetilde{\mathbf{x}}_t(\tau_t)\|_C = O(T^{-1})$. Hence, I have $\max_t \|\mathbf{x}_t\|_C \leq M$.

By Lemma C.3.8 and the definitions of $\mathbf{y}_t$ and $\mathbf{y}_t^c$, I have

$$\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r} |g(\mathbf{y}_t, \boldsymbol{\vartheta}) - g(\mathbf{y}_t^c, \boldsymbol{\vartheta})|\|_1 \leq M\sum_{j=t}^{\infty}\chi_j\|\mathbf{x}_{t-j}\|_C = O\left(\sum_{j=t}^{\infty}\chi_j\right).$$

In addition, by Proposition 4.2.2.1 and $\chi_j = O(j^{-(2+s)})$ for some $s > 0$, I have

$$
\begin{aligned}
& \|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r} |g(\mathbf{y}_t, \boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\vartheta})|\|_1 \leq M\sum_{j=0}^{\infty}\chi_j\|\mathbf{x}_{t-j} - \widetilde{\mathbf{x}}_{t-j}(\tau_t)\|_C \\
\leq\ & M\sum_{j=0}^{\infty}\chi_j\|\mathbf{x}_{t-j} - \widetilde{\mathbf{x}}_{t-j}(\tau_{t-j})\|_C + M\sum_{j=0}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau_t) - \widetilde{\mathbf{x}}_{t-j}(\tau_{t-j})\|_C \\
=\ & O(\sum_{j=0}^{\infty}\chi_j/T) + O(\sum_{j=0}^{\infty}j\chi_j/T) = O(T^{-1}).
\end{aligned}
$$

Hence, I have

$$
\begin{aligned}
& \| \sup_{\tau\in[0,1]} \sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)} |\widetilde{G}_\tau(\boldsymbol{\eta}) - G_\tau^c(\boldsymbol{\eta})| \|_1 \\
\leq\ & M(Th)^{-1}\sum_{t=1}^{T} \sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r} \|g(\widetilde{\mathbf{y}}_t,\boldsymbol{\vartheta}) - g(\mathbf{y}_t^c,\boldsymbol{\vartheta})\|_1 \\
\leq\ & M(Th)^{-1}\sum_{t=1}^{T}\sum_{j=t}^{\infty}\chi_j \leq M(Th)^{-1}\sum_{j=1}^{\infty} j\chi_j = O((Th)^{-1}).
\end{aligned}
$$

The proof of the fourth result is now complete. □

Proof of Lemma C.3.4.

(1). Let $\widetilde{\mathbf{y}}_t^*(\tau)$ be a coupled version of $\widetilde{\mathbf{y}}_t(\tau)$ with $\boldsymbol{\varepsilon}_0$ replaced by $\boldsymbol{\varepsilon}_0^*$. By Lemma C.3.8, I have

$$
\begin{aligned}
\delta_q^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(t) &= \| \sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})| - \sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})| \|_q \\
&\leq \| \sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})| \|_q \\
&\leq M\sum_{j=0}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau) - \widetilde{\mathbf{x}}_{t-j}^*(\tau)\|_{qC} \\
&= M\sum_{j=0}^{t}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j).
\end{aligned}
$$

By Proposition 4.2.1.2 and the conditions on $\alpha_j(\boldsymbol{\theta}(\tau))$ and $\beta_j(\boldsymbol{\theta}(\tau))$ in the body of this lemma, I have $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(j) = O(j^{-(a+s)})$ for some $s>0$. Hence, I have

$$
\begin{aligned}
\sum_{j=0}^{t}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j) &\leq \sum_{j\geq t/2}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j) + \sum_{0\leq j\leq t/2}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j) \\
&\leq (t/2)^{-(a+s)}\sum_{j\geq t/2}\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j) + (t/2)^{-(a+s)}\sum_{0\leq j\leq t/2}\chi_j \\
&= O(t^{-(a+s)}).
\end{aligned}
$$

The proof of the first result of this lemma is now complete.
(2)–(3). Since

$$
\begin{aligned}
|\sup_{u,\boldsymbol{\eta}}|m(\tau,\boldsymbol{\eta},u)| - \sup_{u,\boldsymbol{\eta}}|m^*(\tau,\boldsymbol{\eta},u)|| &\leq \sup_{u,\boldsymbol{\eta}}|m(\tau,\boldsymbol{\eta},u) - m^*(\tau,\boldsymbol{\eta},u)| \\
&\leq M\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})|,
\end{aligned}
$$

the second result follows directly from the first result.

Since $\mathbf{d}(u,\tau) = O(h^2)$ when $|\tau-u|\leq h$, for each element of $\mathbf{m}_t^{(2)}(u,\tau)$, I have

$$
\begin{aligned}
|\sup_{\tau}|m_{t,i}^{(2)}(u,\tau)| - \sup_{\tau}|m_{t,i}^{(2)}(\tau,u)^*|| &\leq \sup_{\tau}|m_{t,i}^{(2)}(\tau,u) - m_{t,i}^{(2)}(\tau,u)^*| \\
&\leq Mh^2\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})|,
\end{aligned}
$$

where $m_{t,i}^{(2)}(\tau,u)$ is yielded by the coupled version.

The proof is now complete. □

Proof of Lemma C.3.5.

(1). Note that

$$\begin{aligned} & g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) - E(g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h)) \\ = & \sum_{l=0}^{\infty} \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h), \end{aligned}$$

in which $\{\mathscr{P}_{t-l}(g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h))\}_{t=1}^T$ is a sequence of martingale differences.

If $1 < q \leq 2$, by the Burkholder's inequality, $|\sum_{i=1}^d a_i|^r \leq \sum_{i=1}^d |a_i|^r$ for $r \in (0, 1]$ and Lemma C.3.4.1, I have

$$\begin{aligned} & \|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q \\ \leq & \sum_{l=0}^{\infty} \left\| \sum_{t=1}^T \frac{1}{Th} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right\|_q \\ \leq & O(1) \sum_{l=0}^{\infty} \left\{ E\left[\sum_{t=1}^T \left(\frac{1}{Th} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^2 \right]^{q/2} \right\}^{1/q} \\ \leq & O(1) \sum_{l=0}^{\infty} \left\{ E\left[\sum_{t=1}^T \left(\frac{1}{Th} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^q \right] \right\}^{1/q} \\ \leq & O(1)(Th)^{-(q-1)/q} \sum_{l=0}^{\infty} \sup_{\tau \in [0,1]} \delta_q^{\sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\vartheta})|}(l) \left(\frac{1}{Th} \sum_{t=1}^T \widehat{K}\left(\frac{\tau_t - \tau}{h}\right)^q \right)^{1/q} \\ = & O((Th)^{-(q-1)/q}). \end{aligned}$$

Similarly, for $q \geq 2$, by the Burkholder's inequality and the Minkowski inequality, I have

$$\begin{aligned} & \|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q \\ = & \left\| \frac{1}{Th} \sum_{t=1}^T \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \sum_{l=0}^{\infty} \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right\|_q \\ \leq & \sum_{l=0}^{\infty} \left\| \sum_{t=1}^T \frac{1}{Th} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right\|_q \\ \leq & O(1) \sum_{l=0}^{\infty} \left\{ E\left[\sum_{t=1}^T \left(\frac{1}{Th} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^2 \right]^{q/2} \right\}^{1/q} \\ \leq & O(1) \sum_{l=0}^{\infty} \left\{ \sum_{t=1}^T \left[E\left(\frac{1}{Th} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^q \right]^{2/q} \right\}^{1/2} \\ = & O(1)(Th)^{-1/2} \sum_{l=0}^{\infty} \sup_{\tau \in [0,1]} \delta_q^{\sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau), \boldsymbol{\vartheta})|}(l) = O((Th)^{-1/2}). \end{aligned}$$

The proof of the first result is now complete.

(2). For any fixed $v > 0$, let $\kappa > 0$ and $\mathbf{E}_T^\kappa(r)$ be a discretization of $\mathbf{E}_T(r)$ such that for each $\boldsymbol{\eta} \in \mathbf{E}_T(r)$ one can find $\boldsymbol{\eta}' \in \mathbf{E}_T^\kappa(r)$ satisfying $|\boldsymbol{\eta} - \boldsymbol{\eta}'| \leq \kappa$. Let $\#\mathbf{E}_T^\kappa(r)$ denote the numbers of sets in $\mathbf{E}_T^{\kappa}(r)$. Write

$$\Pr\left(\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>v\right) \leq \#\mathbf{E}_T^{\kappa}(r)\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}\Pr\Big(|\widetilde{G}_\tau(\boldsymbol{\eta})|>v/2\Big) + \Pr\left(\sup_{|\boldsymbol{\eta}-\boldsymbol{\eta}'|\leq\kappa}|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|>v/2\right).$$

By the Markov inequality, I have

$$\Pr\Big(|\widetilde{G}_\tau(\boldsymbol{\eta})|>v/2\Big)\leq\frac{\|\widetilde{G}\tau(\boldsymbol{\eta})\|_q^q}{(v/2)^q}.$$

Note that $\{\mathscr{P}_{t-j}g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\}_t$ forms a sequence of martingale differences. By the Burkholder's inequality and Lemma C.3.4.1, I have

$$\begin{aligned}
&\|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q\\
\leq\ & (Th)^{-1}\sum_{j=0}^{\infty}\|\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\mathscr{P}_{t-j}g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\|_q\\
\leq\ & (q-1)^{-1}(Th)^{-1}\sum_{j=0}^{\infty}\left(\|\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)^2\mathscr{P}_{t-j}^2g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\|_{q/2}^{q/2}\right)^{1/q}\\
\leq\ & M(Th)^{-(q-1)/q}\sum_{j=0}^{\infty}\sup_{\tau\in[0,1]}\delta_q^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(j)=O((Th)^{-(q-1)/q}),
\end{aligned}$$

which in connection with the fact $\#\mathbf{E}_T^{\kappa}(r)$ is independent of T yields that

$$\#\mathbf{E}_T^{\kappa}(r)\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}\Pr\Big(|\widetilde{G}_\tau(\boldsymbol{\eta})|>v/2\Big)=o(1).$$

In addition, by Lemma C.3.3.1, I have

$$\Pr\left(\sup_{|\boldsymbol{\eta}-\boldsymbol{\eta}'|\leq\kappa}|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|>v/2\right)\leq M\kappa\to 0$$

by choosing κ small enough. Hence, $\Pr\Big(\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>v\Big)\to 0$ as $T\to\infty$.

(3). Let $\beta_T:=(\log T)^{1/2}(Th)^{-1/2}h^{-1/2}$ for short. Let further $\mathbf{E}_{T,\kappa}(r)$ be a discretization of $\mathbf{E}_T(r)$ such that for each $\boldsymbol{\eta}\in\mathbf{E}_T(r)$ one can find $\boldsymbol{\eta}'\in\mathbf{E}_{T,\kappa}(r)$ satisfying $|\boldsymbol{\eta}-\boldsymbol{\eta}'|\leq\kappa_T^{-1}$. Define $\mathscr{T}_{T,\kappa}=\{t/\kappa_T:t=1,2,...,\kappa_T\}$ as a discretization of $[0,1]$. For some constant $M>0$, I have

$$\begin{aligned}
&\Pr\left(\sup_{\tau\in[0,1]}\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>M\beta_T\right)\\
\leq\ &\Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}}\sup_{\boldsymbol{\eta}\in\mathbf{E}_{T,\kappa}(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>\beta_T M/2\right)\\
&+\Pr\left(\sup_{|\tau-\tau'|\leq\kappa_T^{-1},|\boldsymbol{\eta}-\boldsymbol{\eta}'|\leq\kappa_T^{-1}}|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_{\tau'}(\boldsymbol{\eta}')|>\beta_T M/2\right).
\end{aligned}$$

Let $m_t(\tau,\boldsymbol{\eta},u):=\widehat{K}\left(\frac{\tau-u}{h}\right)g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau-u)/h)$, I have $\sup_{\tau\in[0,1]}\sup_{u,\boldsymbol{\eta}}\delta_q^{m(\tau,\boldsymbol{\eta},u)}(j)=$

$O(j^{-(a+s)})$ and $\sup_{\tau\in[0,1]}\delta_q^{\sup_{u,\boldsymbol{\eta}}|m(\tau,\boldsymbol{\eta},u)|}(j) = O(j^{-(a+s)})$ for some $a \geq 3/2$ by Lemma C.3.4.2. Let $\alpha = 1/2$, I have

$$W_{q,\alpha} := \max_{k\geq 0}(k+1)^{\alpha}\sup_{\tau\in[0,1]}\sum_{j=k}^{\infty}\delta_q^{\sup_{u,\boldsymbol{\eta}}|m(\tau,\boldsymbol{\eta},u)|}(j) \leq M\max_k k^{-(a-3/2+s)} < \infty$$

and

$$W_{2,\alpha} := \max_{k\geq 0}(k+1)^{\alpha}\sup_{\tau\in[0,1]}\sup_{u,\boldsymbol{\eta}}\sum_{j=k}^{\infty}\delta_2^{m(\tau,\boldsymbol{\eta},u)}(j) \leq M\max_k k^{-(a-3/2+s)} < \infty.$$

Note that $l = \min\{1, \log(\#\mathbf{E}_{T,\kappa}(r)\times\mathscr{T}_{T,\kappa})\} \leq 3(2d+1)\log(T)$ and $M\beta_T Th = MT^{1/2}(\log T)^{1/2} \geq \sqrt{Tl}W_{2,\alpha} + T^{1/q}l^{3/2}W_{q,\alpha} \geq T^{1/2}(\log T)^{1/2} + T^{1/q}(\log T)^{3/2}$ for some M large enough. By using Theorem 6.2 of Zhang and Wu [2017] (the proof therein also works for the uniform functional dependence measure) with $q > 2$ and $\alpha = 1/2$ to $\{m_t(\tau,\boldsymbol{\eta},\tau_t)\}_{\tau\in\mathscr{T}_{T,\kappa},\boldsymbol{\eta}\in\mathbf{E}_{T,\kappa}(r)}$, I have

$$\begin{aligned}
&\Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}}\sup_{\boldsymbol{\eta}\in\mathbf{E}_{T,\kappa}(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})| > \beta_T M/2\right)\\
\leq\ & \frac{MTl^{q/2}}{(\beta_T Th)^q} + M\exp\left(-\frac{M(\beta_T Th)^2}{T}\right)\\
\leq\ & M\left(T^{-(q-2)/2} + \exp(-\log T)\right) \to 0.
\end{aligned}$$

In addition, by the Markov inequality and Lemma C.3.3.1, I have

$$\Pr\left(\sup_{|\tau-\tau'|\leq\kappa_T^{-1},|\boldsymbol{\eta}-\boldsymbol{\eta}'|\leq\kappa_T^{-1}}|\widetilde{G}_\tau(\boldsymbol{\eta}) - \widetilde{G}_{\tau'}(\boldsymbol{\eta}')| > \beta_T M/2\right) = O(h^{-2}T^{-3}/\beta_T) \to 0.$$

The proof is now complete. □

Proof of Lemma C.3.6.

For notational simplicity, I let $\boldsymbol{\eta}(\tau) = [\boldsymbol{\theta}(\tau), h\boldsymbol{\theta}^{(1)}(\tau)]$ in what follows, and define

$$\begin{aligned}
\boldsymbol{\Gamma}(\tau) \ :=\ & \boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E[\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))]\\
& -\frac{1}{Th}\sum_{t=1}^{T}\widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t)).
\end{aligned}$$

Due to $E(\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))) = 0$ by Lemma C.3.1, I have

$$\begin{aligned}
\boldsymbol{\Gamma}(\tau) \ =\ & \frac{1}{Th}\sum_{t=1}^{T}\widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes[\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau)+\boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau))\\
& -\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))]\\
& -\frac{1}{Th}\sum_{t=1}^{T}\widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes E[\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau)+\boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau))\\
& -\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))].
\end{aligned}$$

By the Mean Value Theorem, I have

$$\boldsymbol{\Gamma}(\tau) = (Th)^{-1}\sum_{t=1}^{T}[\mathbf{M}_t^{(2)}(\tau,\tau_t) - E(\mathbf{M}_t^{(2)}(\tau,\tau_t))],$$

where

$$\mathbf{M}_t^{(2)}(\tau,u) := \widehat{\mathbf{K}}((\tau_t-\tau)/h)\otimes \nabla^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u),\widetilde{\boldsymbol{z}}_{t-1}(u);\boldsymbol{\theta}(\tau)-v\mathbf{r}(u))\mathbf{r}(u) \text{ for some } v\in[0,1],$$
$$\mathbf{r}(u) = \frac{1}{2}\boldsymbol{\theta}^{(2)}(\tau)(u-\tau)^2 + \frac{1}{6}\boldsymbol{\theta}^{(3)}(\overline{\tau})(u-\tau)^3 \text{ with } \overline{\tau} \text{ between } u \text{ and } \tau.$$

I then use a similar argument as in the proof of Lemma C.3.5 to prove

$$\Pr\left(\sup_{\tau\in[0,1]}|\boldsymbol{\Gamma}(\tau)| > M\beta_T h^2\right) \to 0.$$

Define $\kappa_T = T^5$ and $\mathscr{T}_{T,\kappa} = \{t/\kappa_T : t = 1,2,...,\kappa_T\}$ as a discretization of $[0,1]$. For some constant $M>0$, I have

$$\begin{aligned}\Pr\left(\sup_{\tau\in[0,1]}|\boldsymbol{\Gamma}(\tau)| > M\beta_T h^2\right) \;\leq\; & \Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}}|\boldsymbol{\Gamma}(\tau)| > \beta_T h^2 M/2\right)\\ &+\Pr\left(\sup_{|\tau-\tau'|\leq\kappa_T^{-1}}|\boldsymbol{\Gamma}(\tau)-\boldsymbol{\Gamma}(\tau')| > \beta_T h^2 M/2\right).\end{aligned}$$

By Lemma C.3.3.3 and the Markov inequality, I have

$$\Pr\left(\sup_{|\tau-\tau'|\leq\kappa_T^{-1}}|\boldsymbol{\Gamma}(\tau)-\boldsymbol{\Gamma}(\tau')| > \beta_T h^2 M/2\right) = O\left(\frac{h^{-2}\kappa_T^{-1}}{\beta_T h^2 M/2}\right)\to 0.$$

By Lemma C.3.4.3, I have

$$\sup_{u,\tau\in[0,1]}\delta_q^{|\mathbf{M}^{(2)}(\tau,u)|}(j) = O(h^2 j^{-(a+s)}) \quad\text{and}\quad \sup_{\tau\in[0,1]}\delta_q^{\sup_u|\mathbf{M}^{(2)}(\tau,u)|}(j) = O(h^2 j^{-(a+s)})$$

for some $a\geq 3/2$. Let $\alpha=1/2$, I have

$$\widetilde{W}_{q,\alpha} := \max_{k\geq 0}(k+1)^{\alpha}\sup_{\tau\in[0,1]}\sum_{j=k}^{\infty}\delta_q^{\sup_u|\mathbf{M}^{(2)}(\tau,u)|}(j) = O(h^2)$$

and

$$\widetilde{W}_{2,\alpha} := \max_{k\geq 0}(k+1)^{\alpha}\sup_{\tau,u\in[0,1]}\sum_{j=k}^{\infty}\delta_2^{|\mathbf{M}^{(2)}(\tau,u)|}(j) = O(h^2).$$

Using Theorem 6.2 of Zhang and Wu [2017] with $q>2$, $\alpha=1/2$ and $l=\min\{1,\log(\#\mathscr{T}_{T,\kappa})\}\leq 5\log(T)$ to $\{\mathbf{M}_t^{(2)}(\tau,\tau_t)\}_{\tau\in\mathscr{T}_{T,\kappa}}$, I have

$$\begin{aligned}&\Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}}|\boldsymbol{\Gamma}(\tau)| > h^2\beta_T M/2\right)\\ \leq\;& \frac{MTl^{q/2}\widetilde{W}_{q,\alpha}^q}{(\beta_T h^2 Th)^q} + M\exp\left(-\frac{M(\beta_T h^2 Th)^2}{T\widetilde{W}_{2,\alpha}^2}\right)\\ \leq\;& M\left(T^{-(q-2)/2}+\exp(-\log T)\right)\to 0.\end{aligned}$$

The proof is now complete. □

Proof of Lemma C.3.7.

(1). Let $\widehat{\boldsymbol{\eta}}(\tau) := [\widehat{\boldsymbol{\theta}}(\tau)^\top, \widehat{\boldsymbol{\theta}}^\star(\tau)^\top]^\top$ and $\boldsymbol{\eta}(\tau) := [\boldsymbol{\theta}(\tau)^\top, h\boldsymbol{\theta}^{(1)}(\tau)^\top]^\top$. By Lemma C.3.5 and the proof of Theorem 4.2.1, I have

$$\sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\eta}}(\tau) - \boldsymbol{\eta}(\tau)| = o_P(1).$$

By the Taylor expansion, I have

$$\widehat{\boldsymbol{\eta}}(\tau) - \boldsymbol{\eta}(\tau) = -(\widetilde{\boldsymbol{\Sigma}}(\tau) + \mathbf{R}_T(\tau))^{-1}\boldsymbol{\nabla}\mathscr{L}_\tau(\boldsymbol{\eta}(\tau)),$$

where $\mathbf{R}_T(\tau) := \boldsymbol{\nabla}^2\mathscr{L}_\tau(\overline{\boldsymbol{\eta}}) - \widetilde{\boldsymbol{\Sigma}}(\tau)$ and $\widetilde{\boldsymbol{\Sigma}}(\tau) := \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{c}_2 \end{bmatrix} \otimes \boldsymbol{\Sigma}(\tau)$ with $\overline{\boldsymbol{\eta}}$ between $\widehat{\boldsymbol{\eta}}(\tau)$ and $\boldsymbol{\eta}(\tau)$. By Lemma C.3.3 and Lemma C.3.5, I have

$$\begin{aligned}
&\sup_{\tau\in[0,1],\boldsymbol{\eta}\in\mathbf{E}_T(r)} |\boldsymbol{\nabla}^2\mathscr{L}_\tau(\boldsymbol{\eta}) - \widetilde{\boldsymbol{\Sigma}}(\tau,\boldsymbol{\eta})| \\
=& \sup_{\tau\in[0,1],\boldsymbol{\eta}\in\mathbf{E}_T(r)} |\boldsymbol{\nabla}^2\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}) - \widetilde{\boldsymbol{\Sigma}}(\tau,\boldsymbol{\eta})| + O_P((Th)^{-1}) \\
=& \sup_{\tau\in[0,1],\boldsymbol{\eta}\in\mathbf{E}_T(r)} |E[\boldsymbol{\nabla}^2\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta})] - \widetilde{\boldsymbol{\Sigma}}(\tau,\boldsymbol{\eta})| + O_P((Th)^{-1} + \beta_T) \\
=& O_P(\beta_T + (Th)^{-1}) + O(h),
\end{aligned}$$

where $\widetilde{\boldsymbol{\Sigma}}(\tau,\boldsymbol{\eta}) := \int_{-\tau/h}^{(1-\tau)/h} K(u) \begin{bmatrix} 1 & u \\ u & u^2 \end{bmatrix} \otimes \boldsymbol{\Sigma}(\tau, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u)\mathrm{d}u$ and

$$\boldsymbol{\Sigma}(\tau, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u) := E\left(\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u)\right).$$

By Lemma C.3.4.3 and the condition

$$\sup_{\tau\in[0,1]} [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(3+s)})$$

for some $s > 0$, I have $\sup_{\tau\in[0,1]} \delta_q^{\boldsymbol{\nabla}_{\vartheta_j}\ell}(j) = O(j^{-(2+s)})$ for some $s > 0$. By Lemma C.3.10, I have

$$\sup_{\tau\in[0,1]} \left| T^{-1}\sum_{t=1}^T K_h(\tau_t - \tau)\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t), \widetilde{\mathbf{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \right| = O_P((Th)^{-1/2}\log T).$$

Since $E\left(\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\right) = 0$, I further obtain that

$$\begin{aligned}
&\sup_{\tau\in[0,1]} |\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E[\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))]| \\
\leq& \sup_{\tau\in[0,1]} |\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E[\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))] \\
&- \frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t - \tau)/h) \otimes \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))| \\
&+ \sup_{\tau\in[0,1]} \left| T^{-1}\sum_{t=1}^T K_h(\tau_t - \tau)\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t), \widetilde{\mathbf{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \right| \\
=& O_P(h^2\beta_T + (Th)^{-1/2}\log T)
\end{aligned}$$

using Lemma C.3.6.

Hence, by Lemma C.3.3.4, I have

$$\begin{aligned}
&\sup_{\tau\in[0,1]} |\nabla \mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| \\
\leq\ & \sup_{\tau\in[0,1]} |\nabla \mathscr{L}_\tau(\boldsymbol{\eta}(\tau)) - \nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))| + \sup_{\tau\in[0,1]} |\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E(\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))| \\
& + \sup_{\tau\in[0,1]} |E(\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))| \\
=\ & \sup_{\tau\in[0,1]} |E(\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))| + O_P(h^2\beta_T + (Th)^{-1/2}\log T + (Th)^{-1}).
\end{aligned}$$

Since

$$\begin{aligned}
& E\left[\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - \frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t-\tau)/h) \otimes \nabla_{\boldsymbol{\vartheta}} \ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))\right] \\
=\ & -\frac{1}{2}h^2\frac{1}{Th}\sum_{t=1}^T \widehat{\mathbf{K}}((\tau_t-\tau)/h) \otimes \left[E[\nabla^2_{\boldsymbol{\vartheta}} \ell(\widetilde{\boldsymbol{x}}_t(\tau), \widetilde{\boldsymbol{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))]\boldsymbol{\theta}^{(2)}(\tau)\left(\frac{\tau_t-\tau}{h}\right)^2\right] + O(h^3) \\
=\ & \frac{1}{2}h^2\int_{-\tau/h}^{(1-\tau)/h} K(u)[u^2, u^3]^\top \mathrm{d}u \otimes \left(-\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)\right) + O((Th)^{-1} + h^3),
\end{aligned}$$

I have

$$\sup_{\tau\in[0,1]} |\nabla_{\boldsymbol{\eta}_j} \mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| = O_P(h^2\beta_T + (Th)^{-1/2}\log T + (Th)^{-1} + h^{1+j}).$$

for $j=1,2$. Hence, I have $\sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\eta}}_j(\tau) - \boldsymbol{\eta}_j(\tau)| = O_P(h^2\beta_T + (Th)^{-1/2}\log T + (Th)^{-1} + h^{1+j})$ and $\sup_{\tau\in[0,1]} |\mathbf{R}_T(\tau)| = O_P(\beta_T + h + (Th)^{-1})$, where $\widehat{\boldsymbol{\eta}}_j(\tau)$ and $\boldsymbol{\eta}_j(\tau)$ are corresponding to the j^{th} part in their definitions given in the beginning of this proof.

Write

$$\begin{aligned}
& \left|-\widetilde{\boldsymbol{\Sigma}}(\tau)(\widehat{\boldsymbol{\eta}}(\tau) - \boldsymbol{\eta}(\tau)) - \nabla \mathscr{L}_\tau(\boldsymbol{\eta}(\tau))\right| \\
\leq\ & |[\mathbf{I}_{2d} + \widetilde{\boldsymbol{\Sigma}}^{-1}(\tau)\mathbf{R}_T(\tau)]^{-1} - \mathbf{I}_{2d}^{-1}| \cdot |\nabla \mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| \\
\leq\ & |[\mathbf{I}_{2d} + \widetilde{\boldsymbol{\Sigma}}^{-1}(\tau)\mathbf{R}_T(\tau)]^{-1}| \cdot |\widetilde{\boldsymbol{\Sigma}}^{-1}(\tau)\mathbf{R}_T(\tau)| \cdot |\nabla \mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| \\
=\ & O_P(\gamma_T).
\end{aligned}$$

The proof of the first result is now complete.

(2). By Lemma C.3.3 and Lemma C.3.6, I have

$$\begin{aligned}
&\sup_{\tau\in[h,1-h]}\Big|\nabla_{\boldsymbol{\vartheta}}\mathscr{L}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau))+\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\Big|\\
\le&\sup_{\tau\in[h,1-h]}|\nabla_{\boldsymbol{\eta}_1}\mathscr{L}_\tau(\boldsymbol{\eta}(\tau))-\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))|\\
&+\sup_{\tau\in[h,1-h]}|\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))-E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))\\
&-\frac{1}{T}\sum_{t=1}^{T}\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)|\\
&+\sup_{\tau\in[h,1-h]}|E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))+\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)|\\
=&\sup_{\tau\in[h,1-h]}|E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))+\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)|+O_P((Th)^{-1}+\beta_T h^2).
\end{aligned}$$

In addition, by the proof of the first result of this lemma, I have

$$\sup_{\tau\in[h,1-h]}|E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))+\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)|=O(h^3+(Th)^{-1}).$$

The proof is now complete. □

C.5 Computation of the Local Linear ML Estimates

In our numerical studies, I use the function *fminunc* in programming language MATLAB to minimize the negative of log-likelihood function. The initial guess is important when using optimization functions because these optimizers are trying to find a local minimum, i.e. the one closest to the initial guess that can be achieved using derivatives. In this section, I give a possible choice of initial estimates.

I could estimate the coefficients of time-varying VARMA(p,q) model

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

by kernel-weighted least squares method if the lagged $\boldsymbol{\eta}_t$ were given. To obtain a preliminary estimator, I first fit a long VAR model and then use estimated residuals in place of true residuals. Consider the VAR(p_T) model

$$\mathbf{x}_t=\sum_{j=1}^{p_T}\boldsymbol{\Gamma}_j(\tau_t)\mathbf{x}_{t-j}+\boldsymbol{\eta}_t,$$

where p_T is set to be $2(Th)^{1/3}$ in our numerical studies. Then, I compute $\widehat{\boldsymbol{\eta}}_t=\mathbf{x}_t-\sum_{j=1}^{p_T}\widehat{\boldsymbol{\Gamma}}_j(\tau_t)\mathbf{x}_{t-j}$, where $\{\widehat{\boldsymbol{\Gamma}}_j(\tau)\}$ are the local linear least squares estimators. Given $\widehat{\boldsymbol{\eta}}_t$, I am able to estimate $\{\mathbf{A}_j(\tau)\}$, $\{\mathbf{B}_j(\tau)\}$ and $\boldsymbol{\Omega}(\tau)$ as well as their derivatives by local linear least squares method.

In order to achieve identifications, certain restrictions should be imposed on the coefficients of the VARMA model. Suppose there exists a known matrix $\mathbf{R}$ and a vector $\boldsymbol{\gamma}(\tau)$ satisfying

$$\mathrm{vec}(\mathbf{A}_1(\tau),...,\mathbf{A}_p(\tau),\mathbf{B}_1(\tau),...,\mathbf{B}_q(\tau))=\mathbf{R}\boldsymbol{\gamma}(\tau),$$

which follows that

$$\mathbf{x}_t \simeq (\mathbf{z}_{t-1}^\top \otimes \mathbf{I}_m)\mathbf{R}[\boldsymbol{\gamma}(\tau) + \boldsymbol{\gamma}^{(1)}(\tau)(\tau_t - \tau)] + \boldsymbol{\eta}_t,$$

where $\mathbf{z}_t = [\mathbf{x}_t^\top, ..., \mathbf{x}_{t-p+1}^\top, \widehat{\boldsymbol{\eta}}_t^\top, ..., \widehat{\boldsymbol{\eta}}_{t-q}^\top]^\top$. Then the local linear estimator of $(\boldsymbol{\gamma}(\tau), \boldsymbol{\gamma}^{(1)}(\tau))$ is given by

$$\begin{pmatrix} \widehat{\gamma}(\tau) \\ h\widehat{\boldsymbol{\gamma}}^{(1)}(\tau) \end{pmatrix} = \left(\sum_{t=1}^{T} \mathbf{R}^\top \mathbf{Z}_{t-1}^* \mathbf{Z}_{t-1}^{*\top} \mathbf{R} K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{T} \mathbf{R}^\top \mathbf{Z}_{t-1}^* \mathbf{x}_t K_h(\tau_t - \tau),$$

where $\mathbf{Z}_t^* = \mathbf{z}_t \otimes \mathbf{I}_m \otimes [1, \frac{\tau_{t+1}-\tau}{h}]^\top$. Similarly, the local linear estimator of $(\text{vech}(\boldsymbol{\Omega}(\tau)), \text{vech}(\boldsymbol{\Omega}^{(1)}(\tau)))$ is given by

$$\begin{pmatrix} \text{vech}(\widehat{\boldsymbol{\Omega}}(\tau)) \\ h\text{vech}(\widehat{\boldsymbol{\Omega}}^{(1)}(\tau)) \end{pmatrix} = \left(\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t^\top K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{T} \mathbf{Z}_t \text{vech}(\widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top) K_h(\tau_t - \tau),$$

where $\mathbf{Z}_t = [1, \frac{\tau_t - \tau}{h}]^\top \otimes \mathbf{I}_{m(m+1)/2}$.

I next consider the preliminary Estimation of Multivariate GARCH Models. Define $\mathbf{y}_t = \mathbf{x}_t \odot \mathbf{x}_t$ and $\mathbf{v}_t = \mathbf{y}_t - \mathbf{h}_t$. I can rewrite the model as

$$\mathbf{y}_t = \mathbf{c}_0(\tau_t) + \sum_{j=1}^{\max(p,q)} (\mathbf{C}_j(\tau_t) + \mathbf{D}_j(\tau_t))\mathbf{y}_{t-j} + \mathbf{v}_t + \sum_{j=1}^{q} (-\mathbf{D}_j(\tau_t))\mathbf{v}_{t-j}$$

with $E(\mathbf{v}_t|\mathscr{F}_{t-1}) = 0$. Similar to the VARMA model, I am able to estimate $\mathbf{c}_0(\tau)$, $\{\mathbf{C}_j(\tau)\}$ and $\{\mathbf{D}_j(\tau)\}$ as well as their derivatives by local linear least squares method. Consider the VAR(p_T) model

$$\mathbf{y}_t = \sum_{j=1}^{p_T} \boldsymbol{\Phi}_j(\tau_t)\mathbf{y}_{t-j} + \mathbf{v}_t,$$

where p_T is set to be $2(Th)^{1/3}$ in our numerical studies. Then, I compute $\widehat{\mathbf{v}}_t = \mathbf{y}_t - \sum_{j=1}^{p_T} \widehat{\boldsymbol{\Phi}}_j(\tau_t)\mathbf{p}_{t-j}$, $\widehat{\mathbf{h}}_t = \mathbf{y}_t - \widehat{\mathbf{v}}_t$ and $\widehat{\boldsymbol{\eta}}_t = \text{diag}^{-1/2}(\widehat{\mathbf{h}}_t)\mathbf{x}_t$. Hence, the local linear estimator of $(\text{vechl}(\boldsymbol{\Omega}(\tau)), \text{vechl}(\boldsymbol{\Omega}^{(1)}(\tau)))$ is given by

$$\begin{pmatrix} \text{vechl}(\widehat{\boldsymbol{\Omega}}(\tau)) \\ h\text{vechl}(\widehat{\boldsymbol{\Omega}}^{(1)}(\tau)) \end{pmatrix} = \left(\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t^\top K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{T} \mathbf{Z}_t \text{vechl}(\widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top) K_h(\tau_t - \tau),$$

where $\mathbf{Z}_t = [1, \frac{\tau_t - \tau}{h}]^\top \otimes \mathbf{I}_{m(m-1)/2}$ and vechl$(\cdot)$ stacks the lower triangular part of a square matrix excluding the diagonal.