6-2022

Essays on long memory time series and panel models

Shuyao KE
Singapore Management University

ESSAYS ON LONG MEMORY TIME SERIES AND PANEL MODELS

SHUYAO KE

SINGAPORE MANAGEMENT UNIVERSITY

2022

ESSAYS ON LONG MEMORY TIME SERIES AND PANEL MODELS

SHUYAO KE

A DISSERTATION

IN

ECONOMICS

Presented to the Singapore Management University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Economics

2022

Supervisor of Dissertation

PhD in Economics, Programme Director

Essays on Long Memory Time Series and Panel Models

by

Shuyao KE

Submitted to School of Economics in partial fulfillment of the requirements

for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Liangjun Su (Co-Chair/Co-Supervisor)
C.V. Starr Chair Professor of Economics
Tsinghua University

Jun Yu (Co-Chair/Co-Supervisor)
Lee Kong Chian Professor of Economics and Finance
Singapore Management University

Peter C. B. Phillips
Sterling Professor of Economics & Professor of Statistics
Yale University
University of Auckland
University of Southampton
Singapore Management University

Sainan Jin
Professor of Economics
Tsinghua University

Singapore Management University

2022

Abstract

This dissertation studies different long memory models. The first chapter considers a time series regression model where both the regressors and error term are locally stationary long memory processes with time-varying memory parameters, and the regression coefficients are also allowed to be time-varying. We consider a frequency-domain least squares estimator with kernelized discrete Fourier transform and derive its pointwise asymptotic normality and uniform consistency. A specification test on the constancy of coefficients is provided. The second chapter studies a linear regression panel data model with interactive fixed effects where the regressors, factors and idiosyncratic error terms are all stationary but with potential long memory. The setup involves a new factor model formulation for which weakly dependent regressors, factors and innovations are embedded as a special case. Standard methods based on principal component decomposition and least squares estimation, as in Bai (2009), are found to suffer bias correction failure because the order of magnitude of the bias is determined in a complex manner by the memory parameters. To cope with this failure and to provide a simple implementable estimation procedure, frequency domain least squares estimation is proposed. The limit distribution of this frequency domain approach is established and a hybrid selection method is developed to determine the number of factors. The third chapter estimates the memory parameters and test them against spurious long memory of the latent factors in a linear regression model with interactive fixed effects, based on the estimated discrete Fourier transform of the factors. The same asymptotic properties hold as if we use the infeasible true factors for both the memory estimator and the test. This result illustrates how the frequency domain least squares estimator can be applied to further inference other than the regression coefficients.

Contents

Acknowledgment

Firstly, I would like to express my sincere thankfulness to my supervisor, Professor Liangjun Su, for giving me countless instruction and feedback throughout my Ph.D. period. I really appreciate his time, patience and funding to support my Ph.D. research and life. For me, he is not only a mentor, but also a role model because of his admirable passion and immersion into academics. In the future, we may not meet and talk as often as before, but living as a scholar like him will be a life-long motivation to me.

My gratitude extends to Professor Anthony Tay and Nicolas Jacquet as the coordinators of Ph.D. program, for providing me advice about keeping in track of my research. Also I would like to thank Ms Thor Qiu Ling and Amelia Tan for giving me support in my life in Singapore.

I would like to thank my coauthor, Professor Peter C. B. Phillips. Working with him has given me a new perspective on my research because of his insight on the overall theories. I would also like to thank Professor Jun Yu and Professor Jia Li, for giving invaluable comments and advice for my papers. In particular, I am grateful to all my committee members, Professor Liangjun Su, Professor Jun Yu, Professor Peter C. B. Phillips, and Professor Sainan Jin, as they kindly write me a reference letter for my job seeking.

In November 2020, I lost my mother, the only parent of mine, and the pandemic stopped me from seeing her for the last time. The past two years were supposed to be so difficult to me, but it was Professor Liangjun Su who encourages me not to give up. He believes my accomplishment in academics will be the greatest relief to my mother. This dissertation, although imperfect, sends my deep missing to her. May her soul rest in peace.

Chapter 1

Introduction

In the last few decades, long memory has been widely observed and identified as one of the important characteristics of macroeconomic and financial data about their strong level of serial persistence (see e.g. Hassler and Wolters, 1995 and Banerjee and Urga, 2005). In linear regression model with long memory time series data, the classic inference procedures based on the OLS estimator is proved to be invalid by the literature. And an alternative method is to handle the model in frequency domain, which is shown to be powerful in long memory setting and has its own general interest. Hence it is an area with both challenge and importance.

In this dissertation, we extend the literature my analyzing two types of long memory models: (1) time-varying regression, and (2) panel regression with interactive fixed effects. The existence of long memory in these two types of model has been well documented by current empirical studies. For instance, see Coakley et al. (2011), Kuan and Hsu (1998) and Lazarová (2005) for long memory in time series model with potential structural changes or smooth variations; and the nature of how factor in the interactive fixed effects represent the latent aggregate macroeconomic or financial trends (see e.g. Stock and Watson, 1989, 2002) indicates its possibility to be long memory (Granger, 1980). To analyze these two types of model, the current frequency domain least squares estimation is adopted and extended to my model setup, which is new to the literature.

In the second chapter, we propose a linear time-varying regression model, where the time-variation occurs at both the regression coefficients and the memory of all variables involved. Following the idea from nonparametric studies, we kernelized the discrete Fourier transform over frequencies local to zero and conduct the least squares estimation wherein. The pointwise asymptotic normality and uniform consistency of this estimator is proved, and a specification test for the constancy of regression coefficients is also proposed. To improve the finite sample performance, we propose a frequency domain bootstrap and prove its validity. Monte Carlo simulation shows that both our estimator and test perform well in finite sample; and an empirical application in terms of international spillover effects of inflation rate across European countries is conducted.

In the third chapter, we propose a long memory panel linear regression model with interactive fixed effects. The classic least squares estimator based on principal component analysis is examined in this framework and proved to be consistent, which is different from pure time series setup. But the asymptotic distribution is now dependent on the unknown memory parameters that contaminate both the convergence rate and the asymptotic bias. Such issue is difficult to deal with because normally the inference using plug-in estimator of the memory parameters perform poorly in finite sample, and also the terms with memory parameters are complex in their form. To cope with this issue, we extend the frequency domain least squares estimator into the factor model, and prove its consistency and asymptotic normality. Moreover, a self-normalized inference scheme is developed, together with a selection scheme that consistently determines the number of factors. Monte Carlo simulation supports our theory in finite sample, and an empirical illustration of the approach is provided, examining the long-run relationship between debt and economic growth.

In the fourth chapter, we continue with the factor model above, and try to study the estimation and inference of the factor memory parameters and test against the potential spurious long memory. We adopt the classic local Whittle estimator and the test based on the score of its likelihood function, and show that the asymptotics of these objects continue to hold as if we are working with the infeasible true factors instead of their estimators. This illustrates how our frequency domain least squares estimator in factor model is compatible with some popular techniques developed for long memory.

In the fifth chapter we conclude and in the appendix we give the theoretical proofs of our main results and the auxiliary lemmas required.

Chapter 2

Time-Varying Regression with Long Memory

2.1 Introduction

Long memory processes have been widely observed in macroeconomic, financial and other types of data in economy. Different from either $I(0)$ or $I(1)$ processes that characterize stationarity and non-stationarity using an integer order of integration, long memory processes allow this order to lie between 0 and 1, which is normally termed "memory parameter" in the literature. A memory parameter, denoted by d, can be non-integer and thus measures the level of persistence of a time series more generally than the commonly used $I(0)$ and $I(1)$ processes in application. When $d > 0$, the autocovariance function is not absolutely summable over the number of lags, and a long memory time series is still stationary when $d \in \left(0, \frac{1}{2}\right)$ and nonstationary when $d \in \left[\frac{1}{2}, 1\right)$. This property makes some classic inference procedures using long-run variance fail. And in a long range of economic evolution, structural changes can happen frequently, either within a country or internationally, due to technology advances, natural disasters or occurrence of pandemic. Understanding these facts leads to the recognition that methods need to be developed in time series models accounting for both long memory and time-varying nature.

Some current studies have considered coexistence of long memory processes and discrete structural changes, also termed ”structural breaks”. Among empirical studies, Caporale and Gil-Alana (2013) find long memory nature of quarterly real output per capita in the U.S. from 1948 to 2008, and find a change of long memory parameter in the second quarter of 1978. Coakley et al. (2011) find a coexistence of long memory and structural breaks for 16 types of commodity future contracts in 1990-2009. In the meantime, there are theoretical studies dealing with such long memory processes with structural breaks, on which Banerjee and Urga (2005) give a very thorough survey. Among them, Kuan and Hsu (1998) estimate an one-time breakpoint in a process with memory parameter $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ using least-squares estimator. Ray and Tsay (2002) derive a Bayesian method for detecting potentially multiple structural breaks of persistence $d \in \left(0, \frac{1}{2}\right)$ and process level μ in a long memory process. Lazarová (2005) considers a linear regression with long memory regressors and error term, and test the structural changes for all relevant parameters under known breakpoints. For more recent studies, Rachinger (2011) extends the methodology of Bai and Perron (1998) to long memory processes for testing multiple structural breaks in memory parameters, and derive least-square estimation for the breakpoints. Wang et al. (2013) propose a prediction procedure for long memory processes with structural breaks, with either memory parameter d or mean μ changes, or both. And Wenger et al. (2018) consider a test of structural break in the mean of a long memory fractionally integrated series.

In addition to discrete structural change in long memory processes, there are studies involving smooth time-varying memory parameters. For instance, Whitcher and Jensen (2000) estimate time-varying memory parameters based on the time-scaled properties of the wavelet transform. Boutahar et al. (2008) estimate an $ARFIMA(0, d_t, 0)$ process together with time-varying memory parameters d_t as a parametric process. The estimation is based on local Whittle estimator developed by Shimotsu and Phillips (2006); but there is no theory supporting its consistency and inference. Lu and Guegan (2011) try to estimate the time-varying memory parameters in a more generalized k-factor Gegenbauer process, where seasonality of data can be taken care of by this model. The asymptotic behavior is considered under frequency domain rather than time domain. Boubaker (2018) estimates $d(t)$ in an $ARFIMA$ model with Gaussian noise, mimicking the framework of Boutahar et al. (2008) by assuming $d(t)$ evolving as a different parametric process. In a more general framework where no parametric specification is imposed on the process of $d(t)$, Roueff and von Sachs (2011) consider a wavelet-based estimator of memory parameters, while under the same setting Wang (2019) considers Fourier-based Geweke-Porter Hudak (GPH) and local Whittle (LW) estimator.

Moreover, there are studies concerning regression using variables with long memory, as in Lazarová (2005) we mentioned above. This type of regression model is special because if the joint persistence of regressors and error terms are large enough, OLS estimator in time domain will converge slower than the usual $T^{-\frac{1}{2}}$ rate and asymptotic distribution is not Gaussian. This conclusion is pointed out by Robinson and Hidalgo (1997) and proved by Chung (2002). Therefore Nielsen (2005) considers a semiparametric weighted frequency domain least square (WFDLS hereafter) estimator of linear regression model where both regressors and error term are allowed to be stationary long memory processes, and proves its asymptotic normality, a similar FDLS estimator, also called "narrow-band" least squares estimator, is adopted by Christensen and Nielsen (2006) on fractional cointegrated regression model. Nielsen and Frederiksen (2011) extend this method to non-stationary processes. And Shimotsu (2012) develops a local Whittle estimator dealing with fractional cointegrated model, estimating both regression coefficients and memory parameters.

In this paper, we focus on a time-varying linear regression model with long memory regressors and error term that have smooth time-varying memory parameters. This framework has not been covered by current studies yet. To the best of our knowledge, the most relevant study involves Beran (2009) and Preuß (2012), where Beran (2009) considers a time-varying $AR(\infty)$ model and estimates both the regression coefficients and memory parameters using a time domain kernelized maximum likelihood method; while Preuß (2012) gives some important asymptotic results about locally stationary long memory process, and a time-varying regression with polynomial coefficients is considered with Gaussian error term. we complement these two studies by considering regression coefficients without any particular functional form; and in terms of estimator, we adopt the one in Nielsen (2005) using frequency domain weighted least squares estimation, with Fourier transforms being kernelized as in Roueff and von Sachs (2011) and Wang (2019). We prove the pointwise consistency, pointwise asymptotic normality and uniform convergence rate of our estimator; and develop a specification test about the constancy of regression coefficients, which could be the main concern when our estimator is applied in the real data. To overcome the possible poor finite sample performance, we estimate the asymptotic covariance matrix using a frequency domain bootstrap method for locally stationary processes developed by Kreiss and Paparoditis (2015), and prove its validity in our framework. And we conduct the inference using a self-normalized scheme, which avoids estimating memory parameters, which may further affect the finite sample performance.

The rest of this paper is organized as follows. Section 2.2 presents the basic setup of our model. Section 2.3 gives the pointwise limiting distribution and uniform convergence rate of our proposed frequency domain estimator. In Section 2.4 we try to consistently estimate the asymptotic covariance in pointwise limiting distribution using a frequency domain bootstrap, and propose a specification test over the constancy of regression coefficients. Section 2.5 is devoted to Monte Carlo simulation, where we report both the basic pointwise performance of our estimator, and size and power of our bootstrapped test statistic. Section 2.6 illustrates an application of our estimator and bootstrap testing using inflation rate data among European countries, where we evaluate the international inflation spillover over time. Section 2.7 concludes. Proofs and other auxiliary results are in Appendix.

We introduce some notations used in the remainder of this paper. For any complex matrix A, we use $\bar{A}$ and A^* to denote its complex conjugate and conjugate transpose, respectively. Let $|A|^2 = A\overline{A}$ for any complex number A and $|A|^2 = AA^*$ for any complex matrix A. Let $\|A\|$ denote the Frobenius norm if A is a real matrix and $\sqrt{\sum_{i,j} |A_{ij}|^2}$ if A is a complex matrix. Let $f_{X_{aa}}(u,\lambda)$, $f_{\varepsilon}(u,\lambda)$ denote the spectral densities of a-th element of regressor $X_{t,T}$ and error term $\varepsilon_{t,T}$ respectively, and $f_{X_{ab}}(u,\lambda)$ and $f_{X_a\varepsilon}(u,\lambda)$ denote the cross-spectral densities between the a-th and b-th elements of regressor vector $X_{t,T}$ and between the a-th element of regressor vector and error term.

2.2 Model and Estimation

2.2.1 Model

In this paper we consider the following regression model:

$$y_{t,T} = \beta'\left(\frac{t}{T}\right) X_{t,T} + \varepsilon_{t,T},\ t = 1, 2, \ldots, T,$$

where $X_{t,T}$ is a $p \times 1$ vector of regressors, and $\varepsilon_{t,T}$ is the error term. Processes like $X_{t,T}$ and $\varepsilon_{t,T}$ are also termed as "locally stationary long memory processes", which means the memory parameters should satisfy $|d(t)| < \frac{1}{2}$ but are allowed to vary over time. And the smoothness of $d(t)$ makes it possible to approximate the original process with a stationary process local to every t, which is why we call it "locally stationary". Local Stationarity is firstly defined by Dahlhaus (1996) and an extensive review can be seen in Dahlhaus

(2012). We assume both $X_{t,T}$ and $\varepsilon_{t,T}$ are either short- or long-memory,[1] with time-varying memory parameters $d_X\left(\frac{t}{T}\right)$ and $d_\varepsilon\left(\frac{t}{T}\right)$ respectively. Note that $d_X\left(\frac{t}{T}\right) = \left(d_{X_1}\left(\frac{t}{T}\right), \ldots, d_{X_p}\left(\frac{t}{T}\right)\right)'$ is a vector of memory parameters for each argument of $X_{t,T}$. By local stationarity, we require both $d_\varepsilon\left(\frac{t}{T}\right)$ and $d_{X_k}\left(\frac{t}{T}\right)$, $k = 1, \ldots, p$, lie in the interval $\left[0, \frac{1}{2}\right)$. Following Robinson and Hidalgo (1997), Hidalgo and Robinson (2002) and Nielsen (2005), we characterize the persistence of $\{X_{t,T}\}$ and $\{\varepsilon_{t,T}\}$ in frequency domain using their spectral densities, while we do not specify any of their parametric form except for singularity at zero frequency for any $u \in (0,1)$. We allow $d_\varepsilon(u) + d_{X_k}(u) > \frac{1}{2}$, which is termed *collective strong dependence* and makes the usual OLS estimator not asymptotically normal, see the statement in Robinson and Hidalgo (1997) and proofs in Chung (2002) for details.

2.2.2 Estimation

For any $u \in (0,1)$, we propose to estimate $\beta(u)$ by the following minimization problem:

$$\hat{\beta}(u) = \arg\min_{\beta \in \mathbb{R}^p} \sum_{j=1}^{M} \left| \sum_{t=1}^{T} \left(y_{t,T} - \beta' X_{t,T}\right) K_h(t - Tu)\, e^{it\lambda_j} \right|^2 \lambda_j^{2\delta(u)} \qquad (2.2.1)$$

where $K_h(u) = K(u/h)/h$, and $K(\cdot)$ is a kernel function, h is a bandwidth, $\lambda_j = 2\pi j/T$ are Fourier frequencies, and $\lambda_j^{2\delta(u)}$ is a time-varying weight function as given in Nielsen (2005). Note that we allow $\delta(u)$ to be dependent on u as well.

To see the intuition behind the above estimator, we can consider $\beta(u) = \beta$ as a constant. In this special case, Robinson and Hidalgo (1997) and Nielsen

[1]For simplicity, we shall refer to $X_{t,T}$ and $\varepsilon_{t,T}$ as "long memory processes" in this study, although they include the possibility of being short memory.

(2005) give a weighted least squares (WLS) estimator:

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{j=1}^{M} \left| w_y(\lambda_j) - \beta' w_X(\lambda_j) \right|^2 \phi(\lambda_j)$$

where $\phi(\lambda_j)$ is some weight function, and $w_y(\lambda_j)$ and $w_X(\lambda_j)$ are Fourier transforms of y_t and X_t respectively, therefore

$$w_y(\lambda_j) - \beta' w_X(\lambda_j) = \sum_{t=1}^{T} (y_t - \beta' X_t) e^{it\lambda_j}$$

is the Fourier transform of the error term. We can see how the above estimator is equivalent to (2.2.1) if we set $\phi(\lambda_j) = \lambda_j^{2\delta}$ and set the kernel function a constant. When $\beta(u)$ is indeed time-varying, we can follow Roueff and von Sachs (2011) and Wang (2019) and introduce a kernel weight into Fourier transform by assigning more weight on the sample at time t that is closer to Tu. We define the kernelized discrete Fourier transform (DFT hereafter) of $\{y_t\}$ at u as follows:

$$w_y(u, \lambda_j) \equiv \frac{1}{\sqrt{2\pi \sum_{t=1}^{T} K_h^2(t - Tu)}} \sum_{t=1}^{T} y_{t,T} K_h(t - Tu) e^{it\lambda_j}, \qquad (2.2.2)$$

and define $w_X(u, \lambda_j)$ and $w_\varepsilon(u, \lambda_j)$ for $\{X_{t,T}\}$ and $\{\varepsilon_{t,T}\}$ analogously. Then analogous to Nielsen (2005),

$$\hat{\beta}(u) = \left[\sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(w_X(u, \lambda_j) w_X^*(u, \lambda_j) \right) \right]^{-1} \left[\sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(w_X(u, \lambda_j) w_y^*(u, \lambda_j) \right) \right],$$

where we use $\mathrm{Re}(\cdot)$ to denote the real part of a complex matrix. The above estimator is well defined provided $\sum_{j=1}^{M} \mathrm{Re}(w_X(u, \lambda_j) w_X^*(u, \lambda_j))$ is positive definite almost surely. We will study the asymptotic properties of $\hat{\beta}(u)$ in the next section.

2.3 Asymptotic Properties of $\hat{\beta}(u)$

Define the $(p+1)\times 1$ vector $v_{t,T} = (X'_{t,T}, \varepsilon_{t,T})'$. Let C denote a generic constant that could vary across lines. To study the asymptotic distribution of $\beta(u)$ at any fixed $u \in (0,1)$, we make the following assumptions together with some remarks.

Assumption 1. (i) $v_{t,T}$ is generated by a linear process

$$v_{t,T} = \sum_{j=0}^{\infty} A_{t,T}(j)\,\zeta_{t-j},$$

where ζ_t is an i.i.d. sequence such that $\mathbb{E}(\zeta_t \mid \mathcal{F}_{t-1}) = 0$, $\mathbb{E}(\zeta_t\zeta_t' \mid \mathcal{F}_{t-1}) = I_{p+1}$, $\mathbb{E}(\zeta_t\zeta_t' \otimes \zeta_t\zeta_t' \mid \mathcal{F}_{t-1}) = \widetilde{B}$ for a symmetric constant matrix $\widetilde{B}$, and $\mathbb{E}(\zeta_t\zeta_v' \otimes \zeta_s\zeta_w') = I_{p+1}\otimes I_{p+1}$ if $t = v \neq s = w$, and $\mathbb{E}(\zeta_t\zeta_v' \otimes \zeta_s\zeta_w') = \widetilde{C}$ for some sparse constant matrix when $t = s \neq v = w$ and $\mathbb{E}(\zeta_t\zeta_v' \otimes \zeta_s\zeta_w') = \widetilde{P}$ as a permutation matrix when $t = w \neq s = v$, and $\mathbb{E}(\zeta_t\zeta_v' \otimes \zeta_s\zeta_w') = 0$ otherwise; $\{A_{t,T}(j)\}_{j=0}^{\infty}$ is a sequence of $(p+1)\times(p+1)$ coefficient matrices such that

$$A_{t,T}(j) = \begin{pmatrix} A_{X,t,T}(j) \\ A_{\varepsilon,t,T}(j) \end{pmatrix} = \begin{pmatrix} \tilde{A}_{X,t,T}(j) & \mathbf{0}_{p\times 1} \\ \mathbf{0}_{1\times p} & \tilde{A}_{\varepsilon,t,T}(j) \end{pmatrix}$$

and $A_{t,T}(j)$ is square-summable over j for all t and T in the sense that $\sum_{j=0}^{\infty}\|A_{t,T}(j)\|^2 < \infty$.

(ii) There exist matrix-valued functions $A^0(\cdot, j) : (0,1] \to \mathbf{R}^{(p+1)\times(p+1)}$ such that

$$\max_{1\le t\le T}\left\|A^0\left(\frac{t}{T}, j\right) - A_{t,T}(j)\right\| \le \frac{Cl(j)}{T^2} \text{ and} \tag{2.3.1}$$

$$\left\|A^0(u,j) - A^0(v,j)\right\| \le C\,|u-v|\,l(j), \tag{2.3.2}$$

with the function $l(j)$ that is square-summable over j.

Assumption 2. (i) Let $A(u,\lambda) = \sum_{j=0}^{\infty} A^0(u,j)\,e^{ij\lambda}$. The spectral density matrix of $v_{t,T}$

$$f_v(u,\lambda) = \frac{1}{2\pi}|A(u,\lambda)|^2 \tag{2.3.3}$$

satisfies that

$$f_v(u,\lambda) \sim \Lambda(u)^{-1} G\Lambda(u)^{-1} \qquad \text{as} \quad \lambda \to 0^+ \tag{2.3.4}$$

where $\Lambda(u) = Diag\left(\lambda^{d_{X_1}(u)}, \ldots, \lambda^{d_{X_p}(u)}, \lambda^{d_\varepsilon(u)}\right)$, and G is a real, symmetric and positive definite (p.d.) matrix.

(ii) Let $f_{v,ab}(u,\lambda)$ denote the (a,b)-th element of $f_v(u,\lambda)$ for $a,b = 1,\ldots,p+1$, viz., the cross-spectral density between the a-th and b-th elements of $v_{t,T}$ with corresponding memory parameters $d_a(u)$ and $d_b(u)$. $f_{v,ab}(u,\lambda)$ satisfies

$$\left|f_{v,ab}(u,\lambda) - G_{ab}\lambda^{-d_a(u)-d_b(u)}\right| = O\left(\lambda^{\gamma-d_a(u)-d_b(u)}\right) \qquad \text{as} \quad \lambda \to 0^+ \tag{2.3.5}$$

for all $u \in [0,1]$ and some $\gamma \in (1,2]$.

(iii) $\left|\frac{\partial f_{v,ab}(u,\lambda)}{\partial \lambda}\right| = O\left(\lambda^{-1-d_a(u)-d_b(u)}\right)$ as $\lambda \to 0^+$ for any $a,b = 1,\ldots,p+1$.

(iv) For the matrix G, we have that $G_{a,p+1} = G_{p+1,a} = 0$ for all $a = 1,2,\ldots,p$; and the leading $p \times p$ principal submatrix of G, denoted as G_X, is p.d.

Assumption 3. (i) $k(\cdot)$ is a second-order nonnegative kernel that is continuously differentiable, bounded and symmetric, and has compact support $[-1,1]$ such that $\int k(u)\,dx = 1$. $\kappa_{0l} \equiv \int [k(u)]^l du < \infty$ for $l = 2,3,4$.

(ii) $|k(u) - k(v)| \leq \Lambda_1 |u - v|$.

(iii) $k(u)$ monotonically increases when $u \in [-1,0]$.

(iv) $\lim_{T\to\infty}\left|\frac{1}{Th}\sum_{t=1}^{T} k^2\left(\frac{t-Tu}{Th}\right)e^{it\lambda_j}\right| = \theta(u,j)$ for some constant $\theta(u,j) < \infty$ such that there exists a threshold value $\overline{\theta}(u)$, which is dependent on h. And $\theta(u,j) = o\left(\frac{1}{\sqrt{Mh}}\right)$ when $j > \overline{\theta}(u)$ and $\theta(u,j) > 0$ and non-negligible otherwise. And more specifically, $\left|\frac{1}{Th}\sum_{t=1}^{T} k^2\left(\frac{t-Tu}{Th}\right)e^{it\lambda_j}\right| = \theta(u,j) + o\left(\frac{1}{Mh}\right)$

and this convergence rate holds uniformly over j. Also we have

$$\left| \frac{1}{Th} \sum_{t=1}^{T} \left(\frac{t}{T} - u \right) k^2 \left(\frac{t - Tu}{Th} \right) e^{it\lambda_j} \right| = o\left(\theta\left(u, j\right)\right),$$

and

$$\left| \frac{1}{Th} \sum_{t=1}^{T} \left(\frac{t}{T} - u \right)^2 k^2 \left(\frac{t - Tu}{Th} \right) e^{it\lambda_j} \right| = o\left(\theta\left(u, j\right)\right).$$

Assumption 4. As $T \to \infty$, we have

(i) $M \to \infty$, $\frac{M}{T} \to 0$;

(ii) $h = h\left(T\right) \to 0$;

(iii) $\frac{(\log M)^2}{Mh^{2+\delta_*}} \to 0$, where $\delta_* = \frac{1}{1-2\overline{d}_X}$ and $\overline{d}_X = \sup_{p,u} d_{X_p}\left(u\right)$.

Assumption 5. The functional coefficient $\beta\left(u\right)$ is twice continuously differentiable on $(0, 1)$.

Remark 1. Assumption 1 and 2 specify the data generating processes (DGP hereafter) and time-varying spectral densities of locally stationary processes $X_{t,T}$ and $\varepsilon_{t,T}$, which is a multivariate version of the specification in Dahlhaus and Polonik (2009). Note that our definition of a locally stationary process with time-varying memory parameter is equivalent to that in Roueff and von Sachs (2011) and Wang (2019), both of which are modified from Dahlhaus (1996) in the form of spectral representation. Their definition specifies such a process $v_{t,T}$ as

$$v_{t,T} = \int_{-\pi}^{\pi} A_{t,T}\left(\lambda\right) e^{-i\lambda t} dZ\left(\lambda\right),$$

where $\zeta_t = \int_{-\pi}^{\pi} e^{-i\lambda t} dZ\left(\lambda\right)$ is the spectral representation of innovations with $dZ\left(\lambda\right)$ having zero mean and orthogonality between different λ's, and $\mathbb{A}_{t,T}\left(\lambda\right) = \sum_{j=0}^{\infty} A_{t,T}\left(j\right) e^{i\lambda j}$ is the DFT of the sequence $\{A_{t,T}\left(j\right)\}$. Then by substitution we can show that

$$\begin{aligned} v_{t,T} &= \sum_{j=0}^{\infty} A_{t,T}\left(j\right) \zeta_{t-j} = \sum_{j=0}^{\infty} A_{t,T}\left(j\right) \int_{-\pi}^{\pi} e^{-i\lambda(t-j)} dZ\left(\lambda\right) \\ &= \int_{-\pi}^{\pi} \sum_{j=0}^{\infty} A_{t,T}\left(j\right) e^{-i\lambda(t-j)} dZ\left(\lambda\right) = \int_{-\pi}^{\pi} \mathbb{A}_{t,T}\left(\lambda\right) e^{-i\lambda t} dZ\left(\lambda\right). \end{aligned}$$

Note that (2.3.1) in Assumption 1 is a condition similar to but slighly stronger than the corresponding one in Dahlhaus and Polonik (2009). (2.3.2) specifies the smoothness of the function $A^0(\cdot, j)$ over its first argument. Dahlhaus and Polonik (2009) and Roueff and von Sachs (2011) also impose similar conditions to ours.

Remark 2. In Assumption 1 the symmetric matrix $\widetilde{B}$, the sparse matrix $\widetilde{C}$ and permutation matrix $\widetilde{P}$ can be calculated using the assumptions $\mathbb{E}(\zeta_t|\mathcal{F}_{t-1}) = 0$, $\mathbb{E}\left(\zeta_t\zeta_t' \mid \mathcal{F}_{t-1}\right) = I_{p+1}$ the i.i.d. of innovations ζ_t. To be exact,

$$\widetilde{B} = \begin{pmatrix} \boldsymbol{\eta}_{11} & \boldsymbol{s}_{12} & \cdots & \boldsymbol{s}_{1,p+1} \\ \vdots & \ddots & & \vdots \\ \boldsymbol{s}_{p+1,1} & \boldsymbol{s}_{p+1,2} & \cdots & \boldsymbol{\eta}_{p+1,p+1} \end{pmatrix},$$

and

$$\widetilde{C} = \begin{pmatrix} \boldsymbol{e}_{11} & \boldsymbol{e}_{12} & \cdots & \boldsymbol{e}_{1,p+1} \\ \vdots & \ddots & & \vdots \\ \boldsymbol{e}_{p+1,1} & \boldsymbol{e}_{p+1,2} & \cdots & \boldsymbol{e}_{p+1,p+1} \end{pmatrix} \text{ and } \widetilde{P} = \begin{pmatrix} \boldsymbol{e}_{11} & \boldsymbol{e}_{21} & \cdots & \boldsymbol{e}_{p+1,1} \\ \vdots & \ddots & & \vdots \\ \boldsymbol{e}_{1,p+1} & \boldsymbol{e}_{2,p+1} & \cdots & \boldsymbol{e}_{p+1,p+1} \end{pmatrix} \tag{2.3.6}$$

where $\boldsymbol{e}_{ij}$ is a $(p+1)\times(p+1)$ matrix with all elements equal to zero except the (i,j)-th one, and $\boldsymbol{s}_{ij} = \boldsymbol{e}_{ij} + \boldsymbol{e}_{ji} = \boldsymbol{s}_{ji}$, and $\boldsymbol{\eta}_{ii} = \text{diag}(1,\ldots,1,\eta_i,1\ldots,1)$ where $\eta_i = \mathbb{E}(\zeta_{ti}^4)$ with ζ_{ti} the i-th element of ζ_t.[2]

Remark 3. (2.3.5) in Assumption 2 also specifies the rate of convergence of

[2]For example, suppose $p+1=2$, then

$$\widetilde{B} = \begin{pmatrix} \eta_1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & \eta_2 \end{pmatrix},$$

and

$$\widetilde{C} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \text{ and } \widetilde{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

from which you can observe how $\widetilde{P}$ is a permutation matrix in a more obvious manner.

spectral density when frequency shrinks to 0 from the positive side. Define

$$C(u,k) \equiv \int_{-\pi}^{\pi} f(u,\lambda) e^{-ik\lambda} d\lambda = \sum_{j=0}^{\infty} A^0(u,j) A^{0\prime}(u,j+k).$$

$C(u,k)$ is termed as "time-varying covariance" in Dahlhaus and Polonik (2009). This term indicates how we approximate a totally "time-varying" process to a "tangent" process that is indexed by a local parameter u, with the help of smoothness conditions in (2.3.1) and (2.3.2). To see the relation between $\mathrm{Cov}(v_{t,T}, v_{t+k,T})$ and $C(u,k)$, notice that

$$\begin{aligned}
&\mathrm{Cov}(v_{t,T}, v_{t+k,T}) \\
&= \sum_{j=0}^{\infty}\sum_{l=0}^{\infty} A_{t,T}(j)\, \mathbb{E}\left(\zeta_{t-j}\zeta'_{t+k-l}\right) A'_{t+k,T}(l) = \sum_{j=0}^{\infty} A_{t,T}(j) A'_{t+k,T}(j+k) \\
&= \sum_{j=0}^{\infty} A^0\left(\frac{t}{T}, j\right) A^{0\prime}\left(\frac{t+k}{T}, j+k\right) + O\left(T^{-1}\right) \\
&= \sum_{j=0}^{\infty} A^0(u,j) A^{0\prime}(u,j+k) + O\left(T^{-1} + \left|u - \frac{t}{T}\right| + \left|u - \frac{t+k}{T}\right|\right) \\
&= C(u,k) + O\left(T^{-1} + \left|u - \frac{t}{T}\right| + \left|u - \frac{t+k}{T}\right|\right), \qquad (2.3.7)
\end{aligned}$$

where the second equality holds by Assumption 1(i), and the third and fourth ones holds by (2.3.1) and (2.3.2) respectively in Assumption 1(ii). Note that $\left|u - \frac{t}{T}\right|$ and $\left|u - \frac{t+k}{T}\right|$ are both $O(h)$ if they are on the support of our kernel function $K_h(\cdot)$. (2.3.7) is helpful in the asymptotic analysis below.

Remark 4. Here we briefly explain why we impose Assumption 3(iv). This assumption controls the order of the periodogram of our kernel function, which is a very essential part that makes our results different from the time-constant long memory regression model like in Nielsen (2005). To be specific, the convergence rate and asymptotic covariance of $\hat{\beta}(u)$ is partly determined by

$$\frac{1}{\left[\sum_{t=1}^{T} k^2\left(\frac{t-Tu}{Th}\right)\right]^2} \sum_{t=2}^{T}\sum_{s=1}^{t-1} k^2\left(\frac{t-Tu}{Th}\right) k^2\left(\frac{s-Tu}{Th}\right) \cos\left((t-s)\lambda_j\right) \cos\left((t-s)\lambda_k\right). \tag{2.3.8}$$

See (A.2.22) in Supplemental Material for more detail. Suppose we set $k\left(\frac{t-Tu}{Th}\right) = 1$ over $t = 1, \ldots, T$, then (2.3.8) becomes

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} \cos\left((t-s)\lambda_j\right)\cos\left((t-s)\lambda_k\right) \\
&= \frac{1}{2T^2}\sum_{t,s=1}^{T}\left(e^{i(t-s)\lambda_{j+k}} + e^{i(t-s)\lambda_{j-k}}\right) - \frac{1}{2T} \\
&= \frac{1}{2}\cdot \mathbf{1}(j=k) + 0\cdot \mathbf{1}(j\neq k) - \frac{1}{2T},
\end{aligned}$$

because $\sum_{t,s=1}^{T} e^{i(t-s)\lambda_{j+k}} = 0$ holds uniformly over $j, k = 1, \ldots, M$ while $\sum_{t,s=1}^{T} e^{i(t-s)\lambda_{j-k}} = 0$ only when $j \neq k$. This is exactly what Nielsen (2005, pp. 294) obtains in his study, and it explains how serial dependence is weakened if we move data from time domain to frequency domain, and in this case data in different frequencies are asymptotically uncorrelated. However, the asymptotic uncorrelation may not occur when $k\left(\frac{t-Tu}{Th}\right)$ is not a constant over $t = 1, \ldots, T$. To see this, using the same reasoning, (2.3.8) is further given by

$$\frac{1}{\left[\sum_{t=1}^{T} k^2\left(\frac{t-Tu}{Th}\right)\right]^2} \sum_{t=2}^{T}\sum_{s=1}^{t-1} k^2\left(\frac{t-Tu}{Th}\right) k^2\left(\frac{s-Tu}{Th}\right) \cos\left((t-s)\lambda_j\right) \cos\left((t-s)\lambda_k\right) \tag{2.3.9}$$

$$\begin{aligned}
&= \frac{1}{2\left(T^2h^2\kappa_{02} + o(1)\right)} \sum_{t,s=1}^{T} k^2\left(\frac{t-Tu}{Th}\right) k^2\left(\frac{s-Tu}{Th}\right)\left(e^{i(t-s)\lambda_{j+k}} + e^{i(t-s)\lambda_{j-k}}\right) \\
&- \frac{1}{2\left(T^2h^2\kappa_{02} + o(1)\right)} \sum_{t=1}^{T} k^4\left(\frac{t-Tu}{Th}\right),
\end{aligned}$$

where the first term corresponds to periodogram of $k^2\left(\frac{t-Tu}{Th}\right)$ over frequencies λ_{j+k} and λ_{j-k}, while the second term is of smaller order. Suppose we consider an arbitrary frequency λ_q with $q \geq 0$, then a DFT of $k^2\left(\frac{t-Tu}{Th}\right)$ at λ_q normalized by $\frac{1}{Th}$ can be given by

$$\begin{aligned}\frac{1}{Th}\sum_{t=1}^{T}k^2\left(\frac{t-Tu}{Th}\right)e^{it\lambda_q} &= \frac{1}{Th}\sum_{t=1}^{T}k^2\left(\frac{t-Tu}{Th}\right)e^{i2\pi q\left(\frac{t}{T}\right)} \\ &\approx \frac{1}{h}\int_0^1 k^2\left(\frac{x-u}{h}\right)e^{i2\pi qx}dx \\ &= e^{i2\pi qu}\int_{-1}^{1}k^2(v)\,e^{i2\pi(qh)v}dv\end{aligned}$$

using Riemann sum approximation and substitution of variable x. Note that under our setup, $q \in \{0, 1, \ldots, M\}$. When $q = 0$, which corresponds to the case when $j = k$, the above equation is $O(1)$ in norm. When $q \neq 0$ but small enough, we could have qh close to zero at large T. Then by Taylor expansion at $qh = 0$ and Leibniz rule,

$$\begin{aligned}&\int_{-1}^{1}k^2(v)\,e^{i2\pi(qh)v}dv \\ &= \int_{-1}^{1}k^2(v)\,dv + i2\pi qh\int_{-1}^{1}vk^2(v)\,dv - 2\pi^2q^2h^2\int_{-1}^{1}v^2k^2(v)\,dv \\ &- \frac{4i}{3}\pi^3q^3h^3\int_{-1}^{1}v^3k^2(v)\,dv + \frac{2}{3}\pi^4q^4h^4\int_{-1}^{1}v^4k^2(v)\,e^{i2\pi(q^*h^*)v}dv \\ &= \int_{-1}^{1}k^2(v)\,dv - 2\pi^2q^2h^2\int_{-1}^{1}v^2k^2(v)\,dv + O\left(q^4h^4\right),\end{aligned}$$

where in the first equality q^*h^* lies between qh and 0, and all the integrals are finite due to the boundedness of $k(\cdot)$ and finite integral horizon; and in the second equality both the odd order terms are zero because $vk^2(v)$ and $v^3k^2(v)$ are odd functions. The order of the second term above is determined by how qh evolves in asymptotics. For q that is small enough, it is negligible as $T \to \infty$; and suppose $q = q_M \sim Kh^{-1}$ for some constant K close to zero, then the second term above is finite but small as qh is still close to zero.

Then Assumption 3(iv) can be interpreted by specifying $\overline{\theta}(u) \sim Kh^{-1}$ for some constant K as above. Then the first term of (2.3.9) can follow that, for all $j,k\in\{1,\ldots,M\}$,

$$\frac{1}{T^2h^2}\sum_{t,s=1}^{T}k^2\left(\frac{t-Tu}{Th}\right)k^2\left(\frac{s-Tu}{Th}\right)\left(e^{i(t-s)\lambda_{j+k}}+e^{i(t-s)\lambda_{j-k}}\right)$$
$$\to\theta^2\left(u,j+k\right)+\theta^2\left(u,j-k\right).$$

When $j=k$, the leading terms of the above limit is $\theta^2\left(u,2j\right)$ with $2j\leq\overline{\theta}\left(u\right)$; and when $j\neq k$, the above limit has its leading terms with $\min\{j+k,|j-k|\}\leq\overline{\theta}\left(u\right)$, which is equivalent to $|j-k|\leq\overline{\theta}\left(u\right)$. And it is easy to see that in time-constant framework, the only analogous leading term is when $j=k$, while in time-varying framework we also need to take into account when $j\neq k$ but with distance controlled by $\overline{\theta}\left(u\right)\sim Kh^{-1}$. This explain how $\sqrt{h}$ enters the convergence rate of our estimator. And note that when $|j-k|>\overline{\theta}\left(u\right)$, we have $\theta^2\left(u,j+k\right)+\theta^2\left(u,j-k\right)=o\left(\frac{1}{Mh}\right)$. This is a technical assumption to make sure a well-defined asymptotic covarice is obtained.

Remark 5. The other parts of Assumption 3 impose some standard conditions on the kernel function. Assumption 4 imposes some conditions on M, h, and γ. Assumption 5 imposes the smoothness condition on the functional coefficient $\beta\left(\cdot\right)$.

To proceed, we introduce some notations. Let the matrix of time-varying periodogram be

$$I_X\left(u,\lambda\right)\equiv w_X\left(u,\lambda\right)w_X^*\left(u,\lambda\right),$$

and its (a,b)-th element be $I_{X_{ab}}\left(u,\lambda\right)=w_{X_a}\left(u,\lambda\right)w_{X_b}^*\left(u,\lambda\right)$ for any $1\leq a,b\leq p$, where $w_{X_i}\left(u,\lambda\right)$ is the normalized DFT as in (2.2.2) of i-th element $X_{i,t,T}$ of $X_{t,T}$. Denote the corresponding time-varying cross spectral density as $f_{X_{ab}}\left(u,\lambda\right)$. Similar notations apply to both $y_{t,T}$ and $\varepsilon_{t,T}$. Like $A_{t,T}\left(j\right)$, we can partition $A^0(u,j)$ and $A(u,\lambda)$ as follows:

$$A^0(u,j) = \begin{pmatrix} A_X^0(u,j) \\ A_\varepsilon^0(u,j) \end{pmatrix} = \begin{pmatrix} \tilde{A}_X^0(u,j) & \mathbf{0}_{p\times 1} \\ \mathbf{0}_{1\times p} & \tilde{A}_\varepsilon^0(u,j) \end{pmatrix} \text{ and}$$

$$A(u,\lambda) = \begin{pmatrix} A_X(u,\lambda) \\ A_\varepsilon(u,\lambda) \end{pmatrix} = \begin{pmatrix} \tilde{A}_X(u,\lambda) & \mathbf{0}_{p\times 1} \\ \mathbf{0}_{1\times p} & \tilde{A}_\varepsilon(u,\lambda) \end{pmatrix},$$

where, e.g., $A_X(u,\lambda) = \sum_{j=0}^{\infty} A_X(u,j)\, e^{ij\lambda}$ denotes the DFT of $A_X(u,j)$.

For the matrix G defined in Assumption 2, we use $G_{X,ab}$ to denote its (a,b)th element for $a,b = 1,...,p$ and $G_{\varepsilon\varepsilon}$ to denote its $(p+1,p+1)$th element. The first main result in this paper is the asymptotic normality of $\hat{\beta}(u)$. This is formally stated in the following theorem.

Theorem 2.3.1 *Suppose that Assumption 1-5 hold. Then*

$$\Lambda_M^{-1}(u)\, \lambda_M^{d_\varepsilon(u)} \sqrt{Mh} \left(\hat{\beta}(u) - \beta(u) - h\widetilde{\mathbb{B}}_M(u) \right) \xrightarrow{d} \mathcal{N}\left(0, \Gamma^{-1}(u)\, \Omega(u)\, \Gamma^{-1}(u)\right),$$

where $\Lambda_M(u) = diag\Big(\lambda_M^{d_{X_1}(u)}, \ldots, \lambda_M^{d_{X_p}(u)}\Big)$, *and the* (a,b)*-th element of the* $p\times p$ *matrix* Γ *is* $\Gamma_{ab} = \frac{G_{X,ab}}{1-d_{X_a}(u)-d_{X_b}(u)+2\delta(u)}$, *that of* Ω *is*

$$\Omega_{ab} = \Theta^* \left(\frac{G_{X,aa}^{\frac{1}{2}} G_{X,bb}^{\frac{1}{2}} G_{\varepsilon\varepsilon}}{1 - d_{X_a}(u) - d_{X_b}(u) - 2d_\varepsilon(u) + 4\delta(u)} \right)$$

with a finite deterministic function $\Theta^*(\cdot)$. *And the bias term* $\widetilde{\mathbb{B}}_M(u)$ *is given by*

$$\widetilde{\mathbb{B}}_M(u) = h^{-1} \left[\sum_{j=1}^{M} \lambda_j^{2\delta(u)} Re\left(w_X(u,\lambda_j)\, w_X^*(u,\lambda_j)\right) \right]^{-1} \times \left[\sum_{j=1}^{M} \lambda_j^{2\delta(u)} Re\left(w_X(u,\lambda_j)\, \tilde{w}_X^*(u,\lambda_j)\right) \right] \beta^{(1)}(u) = O_p(1),$$

with

$$\widetilde{w}_X(u,\lambda_j) = \frac{1}{c_{Tu}} \sum_{t=1}^{T} \left(\frac{t}{T} - u\right) X_{t,T} K_{h,tu} e^{it\lambda_j}.$$

Remark 6. Theorem 2.3.1 has its convergence rate and asymptotic covariance similar to Theorem 1 of Nielsen (2005), especially they are identical if we set $K_h(t - Tu) = 1$. This also implies how our theorem generalizes the one of Nielsen (2005) by making it time-varying.

In the following, we study the uniform convergence rate of our estimator. To help develop the corresponding results, we require the following extra assumptions.

Assumption 6. $\mathbb{E}\left\|\zeta_t \zeta_t' - I_{p+1}\right\|^s < \infty$ for some $s \geq 2$.

Assumption 7. $\frac{T}{M^\eta h^3 \log M} \to 0$ with $\eta = \frac{C^2}{1+\frac{1}{3}C^2}$for $C \geq 3$ as $T \to \infty$.

Furthermore, we suppose $\delta(u) \geq \underline{\delta}$ for some finite lower bound $\underline{\delta}$, and $\min_p \left|d_{X_p}(u) - d_\varepsilon(u)\right| \geq \triangle\underline{d}$ for some finite lower bound $\triangle\underline{d}$, and both bounds are uniform over $u \in (0,1)$. We define $\mathcal{U} = [h, 1-h]$ so as to avoid the boundary issues of our kernel function when trying to approximate the relevant Riemann sum to an integral. Then the following theorem gives the uniform convergence rate of $\hat{\beta}(u)$.

Theorem 2.3.2 *Suppose Assumptions 1-7 hold,*

$$\sup_{u\in\mathcal{U}} \left\|\hat{\beta}(u) - \beta(u)\right\| = O_p(h) + O_p\left(\lambda_M^{\triangle\underline{d}} \sqrt{\frac{\log M}{Th}}\right). \tag{2.3.10}$$

Theorem 2.3.2 establishes the uniform convergence rate of $\hat{\beta}(u)$ over the compact set $\mathcal{U}$. The two terms on the right hand side (RHS) of (2.3.10) reflect the contribution from the asymptotic bias and variance terms, respectively. Unlike the Nadaraya-Watson (local constant) estimator with weakly dependent observations that exhibits the asymptotic bias of order $O(h^2)$, we can only derive a bias term of order $O(h)$.

2.4 Bootstrap Inference

As we can see in Theorem 2.3.1 the asymptotic covariance of our estimator has no close-form expression, and the "frequency leak" induced by its proof also makes the method of moments over frequency domain infeasible. So to conduct the inference, we propose a bootstrap scheme to estimate the asymptotic covariance. This bootstrap covariance estimator can help self-normalize our estimator, which avoids estimating the time-varying memory parameters that are typically poor in finite sample performance as well. Additionally we consider testing the hypothesis that the functional regression coefficients are constant using our self-normalized estimator.

2.4.1 Bootstrap

Here we propose the bootstrap procedures in frequency domain and prove its validity. Dahlhaus and Janas (1996), among others, propose bootstrapping the integrated periodogram over frequency domain for a stationary linear process. The similar idea is adapted by (Kreiss and Paparoditis, 2015, KP hereafter) for a locally stationary short memory process and by Preuß (2012) for a locally stationary long memory process. Here we follow the basic procedure of KP with some modification and try to generate the bootstrapped version of W_{T,m^*}.

Before we move on to the specific procedure, we introduce the intuition behind our bootstrap method. For a locally stationary long memory linear process $\{X_{t,T}\}_{t=1}^{T}$, consider a DFT over Fourier frequencies $\lambda_j = 2\pi j/T$, $j = 1, \ldots, T$ as we defined before given by

$$\overline{w}_X(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_{t,T} e^{it\lambda_j} \tag{2.4.1}$$

Then $X_{t,T}$ can be recovered asymptotically by an discrete inverse of DFT as

$$X_{t,T} \approx \sqrt{\frac{2\pi}{T}} \sum_{j=1}^{T} \overline{w}_X(\lambda_j) e^{-it\lambda_j}$$

Consider that the linear process $\{X_{t,T}\}_{t=1}^{T}$, as in Assumption 1, is given by $X_{t,T} = \sum_{j=0}^{\infty} A_{t,T}(j) \zeta_{t-j}$. Then the DFT $\overline{w}_X(\lambda_j)$ can be approximated as follows.

$$\begin{aligned}
\overline{w}_X(\lambda_j) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \sum_{k=0}^{\infty} A_{t,T}(k) \zeta_{t-k} e^{it\lambda_j} \\
&= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \sum_{k=0}^{\infty} A_{t,T}(k) e^{ik\lambda_j} \zeta_{t-k} e^{i(t-k)\lambda_j} \\
&\approx \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} A\left(\frac{t}{T}, \lambda_j\right) \zeta_t e^{it\lambda_j} \\
&\approx \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sqrt{f_X\left(\frac{t}{T}, \lambda_j\right)} \zeta_t e^{it\lambda_j}
\end{aligned}$$

by eqn (2.3.1) to (2.3.3) in Assumption 1 and 2. Then $X_{t,T}$ can be further approximated by

$$\begin{aligned}
X_{t,T} &\approx \sqrt{\frac{2\pi}{T}} \sum_{j=1}^{T} \overline{w}_X(\lambda_j) e^{-it\lambda_j} \\
&\approx \sqrt{\frac{2\pi}{T}} \sum_{j=1}^{T} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \sqrt{f_X\left(\frac{s}{T}, \lambda_j\right)} \zeta_t e^{is\lambda_j} e^{-it\lambda_j} \\
&\approx \sqrt{\frac{2\pi}{T}} \sum_{j=1}^{T} \sqrt{\hat{f}_X\left(\frac{t}{T}, \lambda_j\right)} w_\zeta(\lambda_j) e^{-it\lambda_j}
\end{aligned}$$

where in the last step we replace the spectral density $f_X\left(\frac{t}{T}, \lambda_j\right)$ by its estimator $\hat{f}_X\left(\frac{t}{T}, \lambda_j\right)$. We can derive $\hat{f}_X\left(\frac{t}{T}, \lambda_j\right)$ by smoothing the local periodograms, which is defined by

$$\hat{f}_X(u, \lambda) = \frac{1}{2\pi} \sum_{j=1}^{T} \omega\left(\frac{\lambda - \lambda_j}{L}\right) I_{X,L}(u, \lambda_j)$$

where the kernel $\omega\left(\frac{\lambda-\lambda_j}{L}\right)$ is an even function over frequency domain satisfying $\sum_{j=1}^{T}\omega\left(\frac{\lambda-\lambda_j}{L}\right)=1$ and $\sum_{j=1}^{T}\omega^2\left(\frac{\lambda-\lambda_j}{L}\right)\to 0$ as $T\to\infty$; and $I_{X,L}(u,\lambda_j)$ is the localized periodogram defined by

$$I_{X,L}(u,\lambda_j)=\frac{1}{2\pi N}\left|\sum_{p=0}^{N}X_{\lfloor Tu\rfloor+1-\lfloor N/2\rfloor+p,T}e^{it\lambda_j}\right|^2$$

where N is the length of localized window such that $\frac{1}{N}+\frac{N}{T}\to 0$ and we specify $X_{t,T}=0$ if $t\leq 0$ or $t>T$. Note that this is just one of the spectral density estimators, the consistency of this estimator is proved for stationary process in Brockwell and Davis (1991, Chapter 10.4) and extended to locally stationary long memory process in Preuß (2012). Another types of estimator that smooths the pre-periodogram over both time and frequency domain is proposed in Dahlhaus (2012), which is not covered here. In summary, the above set-up for approximating the sample using inverse of DFT fits the one proposed in both KP and Preuß (2012). Next we presents the procedures to generate the bootstrap asymptotic covariance estimator as follows:

1. Obtain the unrestricted residuals $\widehat{\varepsilon}_{t,T}=y_{t,T}-X_{t,T}'\hat{\beta}\left(\frac{t}{T}\right)$ by our estimator proposed at each $t=1,\ldots,T$.

2. Generate pseudo innovations $\{\zeta_t^*\}_{t=1}^{T}$ using $i.i.d.$ standard normal distribution.

3. Generate the DFT for the pseudo innovations above as $\tilde{w}_{\zeta^*}(\lambda_j)=\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\zeta_t^*e^{it\lambda_j}$ at frequencies $\lambda_j=2\pi j/T$, $j=1,\ldots,\lfloor T/2\rfloor$.

4. Calculate the inverse of DFT as $\varepsilon_{t,T}^*=\frac{\sqrt{2\pi T}}{\lfloor T/2\rfloor}\sum_{j=1}^{\lfloor T/2\rfloor}\sqrt{2\pi\hat{f}_{\hat{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)}\tilde{w}_{\zeta^*}(\lambda_j)\,e^{-it\lambda_j}$, where $\hat{f}_{\hat{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)$ is the estimated spectral density for residual $\hat{\varepsilon}_{t,T}$.

5. Generate the bootstrapped dependent variable $y_{t,T}^*=X_{t,T}'\hat{\beta}\left(\frac{t}{T}\right)+\varepsilon_{t,T}^*$ and conduct our estimation that gives $\hat{\beta}_*(u)$ for each $u\in(0,1)$.

6. Repeat step 1-5 for B times, and calculate the bootstrap covariance through Monte Carlo simulation as

$$\widehat{\Sigma}_*(u) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\beta}_*^{(b)}(u) - \hat{\beta}(u)\right)\left(\hat{\beta}_*^{(b)}(u) - \hat{\beta}(u)\right)'$$

where $\hat{\beta}_*^{(b)}(u)$ is the estimator at b-th bootstrap sample. Then a self-normalized estimator is given by $\widehat{\Sigma}_*^{-\frac{1}{2}}(u)\left(\hat{\beta}(u) - \beta(u)\right)$.

For the consistency of our bootstrap method, we need to firstly show that for any $u \in (0,1)$, we have $\lambda_M^{d_\varepsilon(u)}\Lambda_M^{-1}(u)\sqrt{Mh}\left(\hat{\beta}_*(u) - \hat{\beta}(u)\right)$ converge in distribution to the same random object as indicated by Theorem 2.3.1. Note that here the convergence is not conditional on the original sample (X,ε), because rather than resampling the data as ususal bootstrap does, we actually generate a pseudo data that mimic but is independent from the original data. Before presenting our bootstrap CLT, we give some extra assumptions for the spectral density estimator $\hat{f}_{\widehat{\varepsilon}}(\cdot,\cdot)$.

Assumption 8. The spectral density estimator $\hat{f}_{\widehat{\varepsilon}}(\cdot,\cdot)$ satisfies

(i) $\left|\frac{\partial \hat{f}_{\widehat{\varepsilon}}(u,\lambda)}{\partial u}\right| < \infty$ for any $u \in (0,1)$ and $\lambda \in (0,\pi)$;

(ii) $\sup_{u,\lambda}\left|\hat{f}_{\widehat{\varepsilon}}(u,\lambda) - f_\varepsilon(u,\lambda)\right| = o(1)$.

The following theorem gives the CLT of bootstrap regression estimator under the null hypothesis.

Theorem 2.4.1 *Given the conditions for Theorem 2.3.1, 2.3.2 and Assumption 8, as $T \to \infty$,*

$$\lambda_M^{d_\varepsilon(u)}\Lambda_M^{-1}(u)\sqrt{Mh}\left(\hat{\beta}_*(u) - \hat{\beta}(u)\right) \xrightarrow{d^*} \mathcal{N}(0, \Sigma(u))$$

where $\Sigma(u) = \Gamma^{-1}(u)\,\Omega(u)\,\Gamma^{-1}(u)$.

2.4.2 Specification test

Our estimator poses the possibility that the regression coefficients could be time-varying. Therefore testing the constancy of regression coefficients becomes a crucial task in application; and in this subsection we introduce a specification test of this aim. Consider the following null hypothesis:

$$\mathbb{H}_0 : \ \beta(u) = \beta$$

for some constant $p \times 1$ vector β. And the alternative hypothesis is given by $\mathbb{H}_1$, as at least one element of $\beta(u)$ is varying with u. As in Cai and Xiao (2012), we consider a test statistic over the distance between $\hat{\beta}(u_i^*)$ and $\bar{\hat{\beta}}$, in a finite set of points $U^* = \{u_i^*\}_{i=1}^{m^*}$, where $\bar{\hat{\beta}}$ is an estimator of β under $\mathbb{H}_0$. Using the results in Nielsen (2005), we can directly define $\bar{\hat{\beta}}$ as

$$\bar{\hat{\beta}} = \left[\sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(\overline{w}_X(\lambda_j)\,\overline{w}_X^*(\lambda_j)\right)\right]^{-1} \left[\sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(\overline{w}_X(\lambda_j)\,\overline{w}_y^*(\lambda_j)\right)\right], \tag{2.4.2}$$

where $\overline{w}_X(\lambda_j)$ is defined as in (2.4.1). Then considering the finite-dimensional distribution of the difference $\hat{\beta}(u_i^*) - \bar{\hat{\beta}}$ for $i = 1, 2, \ldots, m^*$, we have the following result.

Corollary 2.4.2 *Under the conditions of Theorem 2.3.1 to 2.4.1, and under* $\mathbb{H}_0$, $\left\{\hat{\beta}(u_i^*) - \bar{\hat{\beta}}\right\}_{u_i^* \in U^*}$ *have the following asymptotic joint distribution:*

$$\mathbf{\Phi}_U \sqrt{Mh} \begin{pmatrix} \hat{\beta}(u_1^*) - \bar{\hat{\beta}} \\ \vdots \\ \hat{\beta}(u_{m^*}^*) - \bar{\hat{\beta}} \end{pmatrix} \xrightarrow{d} \mathcal{N} \begin{pmatrix} \Sigma(u_1^*) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma(u_{m^*}^*) \end{pmatrix}$$

where $\mathbf{\Phi}_U$ *is the block diagonal matrix defined by*

$$\mathbf{\Phi}_U = diag\left(\lambda_M^{d_\varepsilon(u_1^*)} \Lambda_M^{-1}(u_1^*), \ldots, \lambda_M^{d_\varepsilon(u_{m^*}^*)} \Lambda_M^{-1}(u_{m^*}^*)\right),$$

and $\Sigma(u_i^*) = \Gamma^{-1}(u_i^*)\,\Omega(u_i^*)\,\Gamma^{-1}(u_i^*)$ *as defined in Theorem 2.3.1.*

We then consider the following two test statistics

$$W_{M,T,m^*} \equiv \max_{1\le i\le m^*} \left\| \hat{\Sigma}_*^{-\frac{1}{2}}(u_i^*)\left(\hat{\beta}(u_i^*) - \bar{\hat{\beta}}\right)\right\|^2, \tag{2.4.3}$$

$$W_{S,T,m^*} \equiv \sum_{i=1}^{m^*} \left\| \hat{\Sigma}_*^{-\frac{1}{2}}(u_i^*)\left(\hat{\beta}(u_i^*) - \bar{\hat{\beta}}\right)\right\|^2. \tag{2.4.4}$$

For this test statistic, we have the following asymptotic null distribution.

Corollary 2.4.3 *Under the conditions of Theorem 2.3.1 to Theorem 2.4.1 and corollary 2.4.2, and under* $\mathbb{H}_0$*,* W_{M,T,m^*} *and* W_{S,T,m^*} *have the following asymptotic null distribution:*

$$W_{M,T,m^*} \xrightarrow{d} \max_{1\le i\le m^*} \chi_i^2(p), \quad \text{and } W_{S,T,m^*} \xrightarrow{d} \sum_{i=1}^{m^*} \chi_i^2(p)$$

where $\chi_1^2(p), \ldots, \chi_{m^*}^2(p)$ *are independent chi-square distribution with* p *degrees of freedom.*

2.5 Monte Carlo Simulation

In the Monte Carlo simulation experiment of our estimator and test statistic, we consider the following data generating process:

$$y_{t,T} = X_{t,T}'\boldsymbol{\beta}\left(\frac{t}{T}\right) + \varepsilon_{t,T} \quad t = 1, 2, \ldots, T$$

where without loss of generality, $X_{t,T}$ is set to be a 2×1 vector. Both $X_{t,T}$ and $\varepsilon_{t,T}$ are generated by time-varying ARFIMA process, and specifically $X_{t,T}$ is generated by $ARFIMA\left(0, d_X\left(\frac{t}{T}\right), 0\right)$, and $\varepsilon_{t,T}$ by $ARFIMA\left(0, d_\varepsilon\left(\frac{t}{T}\right), 0\right)$, where these time-varying memory parameters are defined by

$$d_{X_1}(u) = d_{X_2}(u) = (1 - \cos(\pi u/3)), \quad u \in [0,1]$$

and

$$d_{\varepsilon}(u) = (1 - \cos(\pi u/5)), \quad u \in [0,1].$$

And the time-varying regression coefficient vector $\boldsymbol{\beta}(u)$ is given by $\beta_1(u) = 2\sin(2\pi u)$ and $\beta_2(u) = 2\cos(2\pi u)$. In terms of orders of the size of frequency M and bandwidth h, we specify $M = \left\lfloor T^{\frac{4}{5}} \right\rfloor$ and $h = 0.1 S_x T^{-\frac{1}{5}}$ where S_x is the average of standard deviation of each element of $X_{t,T}$. In the following we report the results of averaged bias, root mean-squared error, and standard error of the estimator $\hat{\boldsymbol{\beta}}(u)$ at $u = 0.2$, 0.4, 0.6 and 0.8. The results are averaged over $R = 200$ repetitions, with sample sizes 100, 200 and 400. In detail, the averaged bias is defined by $\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\right)$, mean-squared error by $\left[\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\right)^2\right]^{\frac{1}{2}}$, and standard error by $\left[\frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\beta}(u) - \overline{\hat{\beta}}(u)\right)^2\right]^{\frac{1}{2}}$. For weight function $\delta(u)$, we set $\delta(u) = 3.5$, 4.5 and 5.5 respectively. See Table 1 for the above results. Also we present the graphs in Figure 1 comparing the curves of true parameters $\boldsymbol{\beta}(u)$ with the averaged estimator (WFDLS) $\hat{\boldsymbol{\beta}}(u)$ at $T = 400$ and $\delta(u) = 5.5$.

Then in terms of the specification test, we specify the following null and alternative hypothesis.[3]

$$\mathbb{H}_0 : \boldsymbol{\beta}(u) = \boldsymbol{\beta} = \left(\sqrt{2}, 2\sqrt{2}\right)',$$

$$\mathbb{H}_1 : \boldsymbol{\beta}(u) = (2\sin(2\pi u), 2\cos(2\pi u))'.$$

The estimator of β under the null is given by (2.4.2) and test statistics given by (2.4.3) and (2.4.4). The collection of points $\{u_i^*\}_{i=1}^{m^*}$ are selected as (almost) equally-spaced in the set $\left\{\frac{t}{T}\right\}_{t=1}^{T}$. By our theory, under the null the above test statistics converge in distribution to $\max_{1\le i\le m^*}\chi_i^2(2)$ and $\sum_{i=1}^{m^*}\chi_i^2(2)$ respectively with m^* independent chi-square distribution with degree of freedom 2, and we therefore can obtain the asymptotic 95% critical value of these two test

[3]The null and alternative are set this way so that the signal-to-noise ratio of our data generating process is restricted around 4.

statistics, denote by $C_{M,0.95}$ and $C_{S,0.95}$. For bootstrapping, we set bootstrap sample size as $B = 300$. In the following we report the size and power under the case that $m^* = 2$, 4, 8 and 12 in Table 2, where we focus on the case $\delta(u) = 5.5$. From Table 1 we can see that our estimator converges properly to true coefficients and Figure 1 shows how the estimated curve is close to its true values as well. From Table 2 the size control and power both perform well using our bootstrap covariance estimator.

Table 1: Results of WFDLS estimates of $\beta(u)$

			$\beta_1(u)$				$\beta_2(u)$			
			$u=0.2$	$u=0.4$	$u=0.6$	$u=0.8$	$u=0.2$	$u=0.4$	$u=0.6$	$u=0.8$
$\delta(u)=3.5$	RMSE	$T=100$	0.973	0.923	0.957	1.032	1.192	1.085	1.050	0.972
		$T=200$	0.799	0.771	0.884	0.843	0.885	0.861	0.825	0.742
		$T=400$	0.626	0.632	0.644	0.734	0.632	0.623	0.662	0.681
	STD	$T=100$	0.975	0.924	0.956	1.032	1.194	1.086	1.042	0.971
		$T=200$	0.800	0.772	0.885	0.844	0.885	0.861	0.826	0.736
		$T=400$	0.627	0.633	0.644	0.735	0.632	0.624	0.663	0.679
$\delta(u)=4.5$	RMSE	$T=100$	1.130	0.986	1.072	1.133	1.360	1.249	1.177	1.074
		$T=200$	0.877	0.865	0.974	0.930	0.991	0.949	0.925	0.800
		$T=400$	0.690	0.692	0.745	0.809	0.707	0.687	0.737	0.745
	STD	$T=100$	1.131	0.987	1.073	1.133	1.362	1.250	1.168	1.072
		$T=200$	0.878	0.866	0.974	0.931	0.991	0.949	0.925	0.794
		$T=400$	0.691	0.693	0.744	0.811	0.707	0.688	0.738	0.743
$\delta(u)=5.5$	RMSE	$T=100$	1.287	1.044	1.207	1.250	1.501	1.407	1.298	1.198
		$T=200$	0.941	0.977	1.040	1.034	1.099	1.036	1.011	0.861
		$T=400$	0.758	0.746	0.829	0.872	0.787	0.747	0.806	0.803
	STD	$T=100$	1.289	1.044	1.209	1.250	1.503	1.408	1.290	1.196
		$T=200$	0.942	0.978	1.040	1.035	1.098	1.036	1.012	0.856
		$T=400$	0.759	0.747	0.828	0.873	0.787	0.748	0.807	0.800

Table 2: Size and Power of Specification Test

		Size (%)				Power (%)			
	m^*	2	4	8	12	2	4	8	12
W_{M,T,m^*}	$T=100$	13.5	16.5	28.0	36.0	80.0	93.5	99.0	99.5
	$T=200$	4.5	4.5	3.0	9.0	77.5	96.5	99.5	100
	$T=400$	3.5	5.0	5.5	5.5	89.5	98.0	100	100
W_{S,T,m^*}	$T=100$	13.0	15.5	31.5	39.5	81.5	96.0	99.5	99.5
	$T=200$	5.0	6.5	5.5	5.5	80.0	98.0	100	100
	$T=400$	4.5	6.0	2.5	6.5	93.0	99.0	100	100

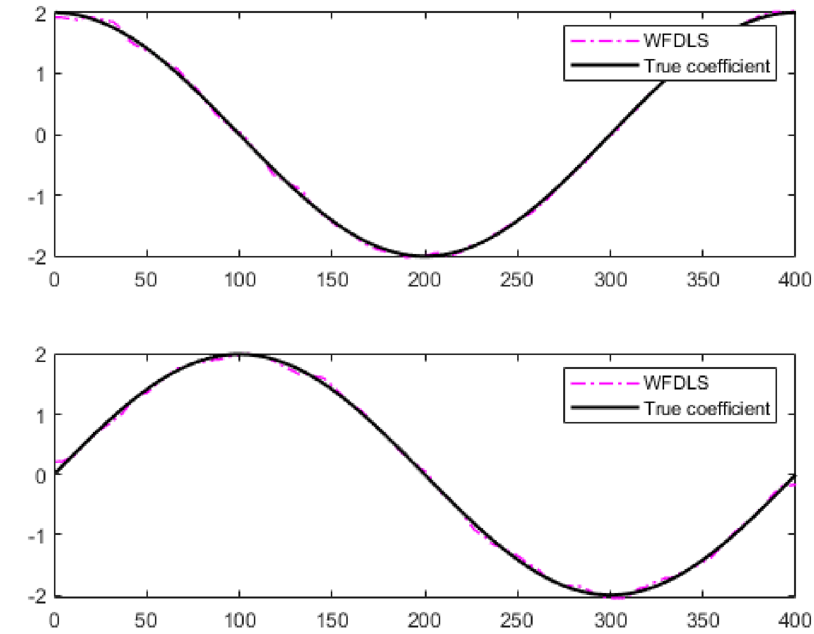

Figure 1: WFDLS estimator against true regression coefficients

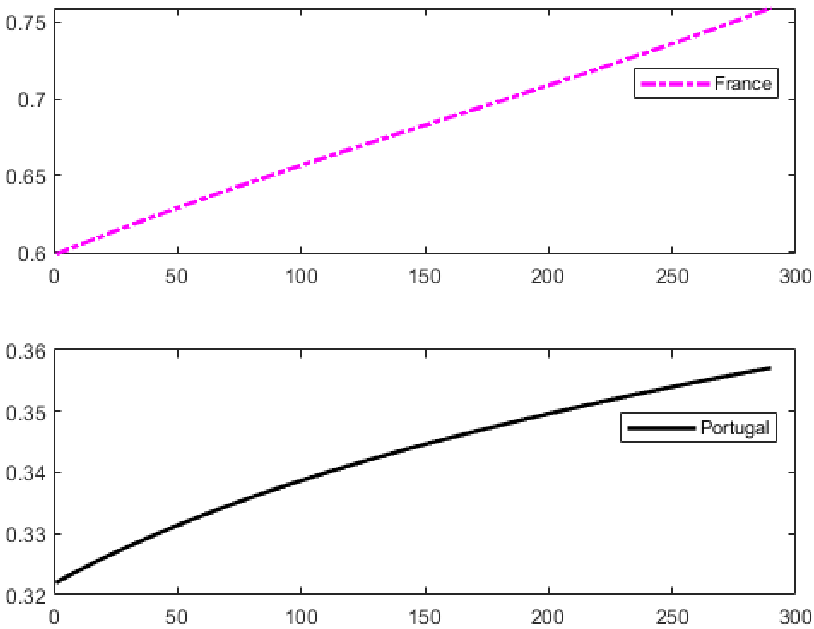

Figure 2: Spillover effects from France and Portugal

2.6 Application

In this section, we present an empirical application in terms of the international inflation spillover effect. The estimation and test for constancy of the effect are based on the harmonized index of consumer prices (HICP hereafter) data collect from Eurostat, an thorough online economic database of EU countries. The HICP are monthly data from January 1996 to March 2020; and different from pure CPI data, HICP is comparable across different countries and thus suitable for our analysis. The nature of long memory for inflation rate is confirmed by Hassler and Wolters (1995) and Doornik and Ooms (2004), among others. And the international inflation spillover has its evidence found by Neely and Rapach (2011), Mumtaz and Surico (2012), Altansukh et al. (2017) and Kang et al. (2019), among others. They found that there is an "interdependence" of inflation across countries, especially for countries that are economically or geographically connected to each other. One of the methods to evaluate inflation spillover effect is to characterize it by a constant regression coefficient along the inflation rates of different countries, see, for example, Nielsen and Frederiksen (2011). However, as an indicator of long-run equilibrium, the time-varying nature of inflation spillover should be considered, which is still empty in current literature. Therefore we try to estimate a time-varying inflation spillover effect to Spain from two other countries: France and Portugal, which are geographically connected to Spain on the ground. This can be illustrated by the following model:

$$InflaSP_{t,T} = \beta_1\left(\frac{t}{T}\right) \cdot InflaFR_{t,T} + \beta_2\left(\frac{t}{T}\right) \cdot InflaPO_{t,T} + \varepsilon_{t,T}$$

where $InflaSP_{t,T}$, $InflaFR_{t,T}$ and $InflaPO_{t,T}$ are inflation rates of Spain, France and Protugal respectively, which are calculated by percentage change of HICP data over every two consecutive periods, with T = 290. Descriptive statistics of these three variables are presented in Table 3. For estimation of

$\beta_1\left(\frac{t}{T}\right)$ and $\beta_2\left(\frac{t}{T}\right)$, we specify the size of frequency $M = \left\lfloor T^{\frac{4}{5}} \right\rfloor$ and bandwidth $h = 5.5 S_x T^{-\frac{1}{5}}$ over all its usage, which is the same as in simulation, and the graphs of $\hat{\beta}_1\left(\frac{t}{T}\right)$ and $\hat{\beta}_2\left(\frac{t}{T}\right)$ are presented in Figure 2.

Next we consider test of constancy of our functional coefficients $\beta_1\left(\frac{t}{T}\right)$ and $\beta_2\left(\frac{t}{T}\right)$, with the following null hypothesis:

$$\mathbb{H}_0 : \begin{pmatrix} \beta_1(u) \\ \beta_2(u) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

for some constants β_1 and β_2. We conduct the test using number of points $m^* = 2$, 4, 8 and 12; and present the value of W_{M,T,m^*}, W_{S,T,m^*} and corresponding 95% critical value as indicated by self-normalization in our bootstrap procedures, denoted by p_M and p_S. The results are presented in Table 4. From Figure 2 we can see the spillover effects of inflation rate in Spain from France and Portugal both grow stably overtime, while the effect and its growth are more prominent from France than from Portugal. From Table 4 we can see that for most of the scenario we reject our null hypothesis, which gives a valid support of time-varying nature of the inflation rate spillover effects.

Table 3: Descriptive Statistics in Application

	Max	Min	Median	Mean	Std. Err
$InflaSP$	2.391	-2.458	0.194	0.170	0.701
$InflaFR$	1.147	-1.122	0.127	0.122	0.329
$InflaPO$	2.239	-1.719	0.122	0.159	0.588

Table 4: Specification Test in Application

m^*	2	4	8	10	12
W_{M,T,m^*}	7.976	15.180	19.454	15.513	21.082
	(7.483)	(8.657)	(10.055)	(10.601)	(10.855)
W_{S,T,m^*}	8.190	23.228	51.171	49.026	82.172
	(9.560)	(15.694)	(26.365)	(31.444)	(36.377)

2.7 Conclusion

In this paper we consider a weighted frequency-domain least square estimation using kernelized discrete Fourier transform of the data, which consistently estimate the time-varying functional regression coefficients. We establish the pointwise asymptotic normality of our estimator, and derive its uniform convergence rate. Also to handle the potential bad finite sample performance, we adopt a bootstrap method that takes use of pseudo innovations in our linear processes and estimation of spectral density for the residuals. We proves its validity in our framework, and apply it to a test for constancy of our regression coefficients. An application to estimation of international inflation spillover effect, using inflation rate data of three European countries, is conducted; and our bootstrap test reject the null hypothesis that these effects are constant over time.

There are several possible extensions. First, in terms of our test for constancy, one may consider a test statistic that evaluates uniformly on the interval $[0, 1]$, this method shall bring to a better finite sample performance than the one we consider now. Second, one may consider another bootstrap method with less tuning parameters considered, as ours now need to involve the bandwidths used for kernelized DFT, for estimating spectral density of the residuals, and for local time window to construct the local periodogram. And as we can see in application, selection of these bandwidth could be burdensome somehow, so at least a data-generated bandwidth selector should be developed. Third, one may extend the current regression framework to one involving endogeneity, especially autoregression; or consider a panel data where fixed effect is considered. So far the frequency domain estimation has not been widely used in panel data with long memory. We leave this as potential improvement of this paper or new topics in the future.

Chapter 3

Unified Factor Model Estimation and Inference under Short and Long Memory

3.1 Introduction

For the past two decades, linear panel regression models with interactive fixed effects have been extensively studied in econometrics and applied in a wide variety of contexts where large datasets have become available in the social and business sciences. These models allow for cross section dependence of an *a priori* unknown form through the use of latent factors that evolve over time with individual loadings that determine the strength of the interactions and temporal dependencies in the panel. The abbreviation "factor model" is used here to represent this general class of panel factor model.

For these factor models to be useful in applied research, it is important that the time series properties of the regressors, factors and innovations in the generating mechanism match those that are present in or implied by the observed data. In practical work it is often convenient to transform dependent variables and regressors to stationarity so that the working model involves a panel of stationary time series. But such transformations do not eliminate the possibility of stationary long range dependence or long memory in the data. To address the complications that can arise through the presence of long memory, the present paper studies a panel linear regression model with interactive fixed effects wherein regressors, factors and idiosyncratic error terms are all stationary but may be driven by long memory processes. The model setup therefore involves a long memory formulation of the factor model in which short memory regressors, factors and innovations are embedded as a special case.

Panel factor model regressions are commonly used in modeling heterogeneous individual behavior that relates to consumption, investment, inflation rates, stock returns, volatility and various other economic and financial indicators. Empirical evidence of long memory has been noted in many of these indicators, implying autocorrelation structures that differ from short memory stationary $I(0)$ processes. For instance, Hassler and Wolters (1995) examined monthly inflation rates for five developed countries and confirmed the presence of long memory in the time series. Similar empirical evidence was found by Caporale and Gil-Alana (2007) for the US unemployment rate, by Gil-Alana and Robinson (2001) for domestic income and consumption in the UK and Japan, and by Ding et al. (1993), Andersen et al. (2001) and Andersen et al. (2003) for stock returns, realized stock volatility and realized exchange rate volatility, respectively.

In applied macroeconomic research, factor modeling is frequently employed to capture the effects of latent aggregate macroeconomic or financial trends, e.g., Stock and Watson (1989, 2002). It is also well known that cross section aggregation of time series can lead to the presence of long memory, as shown by Granger (1980) and studied in economic and financial data by Chambers (1998), Pesaran and Chudik (2014), and Michelacci and Zaffaroni (2000). Long range dependence features in the data and processes like aggregation that underlie much data collection both motivate studying the impact of such dependence on current methods of panel factor modeling and the development of new methods to address the existence of long memory in the data.

The present paper undertakes this investigation and development. In particular, we study estimation, inference, and associated asymptotics for the fitted coefficients in a linear panel regression model with interactive fixed effects with potential long memory regressors, factors and idiosyncratic errors. The starting point of the analysis is standard principal components least squares estimation (Bai (2009)) and its asymptotic performance under long memory. The results of this analysis reveal that, when the joint memory properties of variables in the model is strong enough, least squares estimation produces non-negligible asymptotic bias which is not resolved either by analytical correction, as suggested in Bai (2009), or by the standard half-panel jackknife methods, proposed in Fernández-Val and Weidner (2016). The reason for this breakdown is that the order of magnitude of the bias depends critically on the memory parameters, as does the convergence rate of the least squares regression coefficient estimator. Different from pure time series long memory regression, the least squares estimator of factor model still obtains an asymptotic normal distribution due to the commonly assumed weak dependence over cross-sectional units, and the condition that the number of cross-sectional units goes to infinity in a comparable order with the number of time periods. Moreover, the convergence rate and bias order can vary across the setting in which regressors and factors are mean zero or mean non-zero, and their joint memory together with idiosyncratic error term.

The above issues substantially complicate successful practical implementation of least squares regression. To resolve these difficulties, the present paper proposes an alternative approach to time domain regression by using frequency domain regression methods that have a long history of successful use in time series regression. These methods originated in the pathbreaking studies of Hannan (1963, 1970) on spectral regression, were further developed for principal components by Brillinger (2001), for trending time series regression (Phillips, 1991; Corbae et al., 2002), with higher order approximations in time series regression (Xiao and Phillips, 1998), and recently have been implemented in long memory time series regressions (e.g. Nielsen, 2005) and in time-dependent frequency domain principal components modeling (Ombao and Ho, 2006). In the factor model context, the procedure follows the usual approach of transforming the model by taking discrete Fourier transforms at the Fourier frequencies, and performing principal components analysis (PCA) in the frequency domain on the system and least squares spectral regression estimation. The combination of PCA and spectral least squares regression yields consistent coefficient estimation and asymptotic normality under general conditions. The asymptotic bias involved in frequency domain estimation can be corrected and the asymptotic variance matrix can be estimated using a frequency domain analytic analogue of the formula used in Bai (2009). Inference is conducted using a self-normalized statistic for which there is no need for separate estimation of the memory parameters that occur in the asymptotic bias and covariance matrix, a feature that simplifies implementation and improves finite sample performance.

This study contributes to the current literature in two ways. First, we extend the range of application of the factor model developed in Bai and Ng (2002), Bai (2003, 2009) and Moon and Weidner (2015), by accounting for long memory and nesting short memory applications as a special case. Second, we contribute to the literature of time series long memory modeling, studied by Robinson and Hidalgo (1997), Marinucci and Robinson (2001), Nielsen (2005) and Christensen and Nielsen (2006) among others, by extending spectral regression estimation and inference to the panel factor model. Specifically, the approach developed extends narrow-band spectral estimation in time series regression to the panel factor model, showing that asymptotic normality in this context holds irrespective of the joint memory of the variables, a result that arises from cross section aggregation and contrasts with time series least squares regression for which the limit theory is known to be non-normal when the sum of the memory parameters of the regressors and the errors exceeds 0.5 (Chung, 2002).

Other recent work has considered the impact of long memory time series in panel data modeling, notably Ergemen and Velasco (2017), Ergemen (2019) and Cheung (2021). Ergemen and Velasco (2017) and Ergemen (2019) study a fractionally integrated factor model where the factors are removed by the methods introduced by Pesaran (2006), projecting the regression on a fractionally integrated cross-sectional average. Our study differs from these papers by using a semiparametric formulation of the long memory components and our approach relies employs PCA in the frequency domain to estimate the discrete Fourier transforms of the factors. Similar to our approach but working in a pure factor model, Cheung (2021) seeks to estimate the memory parameters of the latent factors by PCA. Cheung (2021) focuses on a fully parametric fractional integrated process and deals with possible nonstationarity, a feature that our study does not include. On the other hand, our study complements the results of Cheung (2021) by providing a limit theory for estimation of and inference concerning the coefficients in a panel linear regression model with latent factors.

The rest of this paper is organized as follows. Section 3.2 introduces the factor model with possible long memory in the component variables. Section 3.3 develops the asymptotics of least squares estimation in time domain, as in Bai (2009) but allowing for stationary long memory. Section 3.4 provides the corresponding analysis in the frequency domain. Section 3.5 proposes an estimate of the true number of factors that is based on the eigenvalue-ratio method developed by Ahn and Horenstein (2013), establishing its consistency under certain conditions. Section 3.6 reports the results of Monte Carlo simulations that explore the finite sample performance of panel least squares estimation in both time and frequency domain formulations, demonstrating some of the difficulties that are involved in time domain estimation. Section 3.7 provides an empirical application of our panel frequency domain procedures to investigate the long-run relationship between GDP and private debt levels for a panel of 21 countries. Section 3.8 concludes. Proofs and other auxiliary technical results are given in the appendix.

The following notation is adopted. For an arbitrary $m \times n$ matrix A, its transpose is denoted by A'; and its conjugate and conjugate transpose are denoted by $\overline{A}$ and A^* respectively if A is complex; moreover its Frobenius norm is $\|A\| = \sqrt{\text{tr}\,(A'A)}$ if A is real, or $\|A\| = \sqrt{\text{tr}\,(A^*A)}$ if A is complex. The spectral norm of A is $\|A\|_{\text{sp}} = \sqrt{\mu_1\,(A'A)}$, when A is real, and $\|A\|_{\text{sp}} = \sqrt{\mu_1\,(A^*A)}$, when A is complex, where $\mu_1\,(\cdot)$ denotes the largest eigenvalue of the Hermitian matrix argument. Let $\mathbb{I}_R$ denote an R-dimensional identity matrix. For any two matrix-valued sequences A_j and B_j of the same dimension, $A_j \sim B_j$ is defined by $\frac{A_{j,(m,n)}}{B_{j,(m,n)}} \to 1$ as $j \to \infty$ for each of its (m,n)-th elements.

3.2 Model

In this paper we consider the data generating process that is given by following linear panel regression model

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + \varepsilon_{it}, \; i = 1, \ldots, N, \; t = 1, \ldots, T, \tag{3.2.1}$$

with a P-vector of regressors X_{it}, common regression coefficients β, and an R-vector of latent factors F_t with factor loading vectors λ_i, and idiosyncratic errors ε_{it}. This study allows X_{it}, F_t and ε_{it} to be stationary long memory time series with respective memory parameter vectors given by $d_X = (d_{X_1}, \ldots, d_{X_P})'$ and $d_F = (d_{F_1}, \ldots, d_{F_R})'$, and memory parameter of ε_{it} given by a scalar d_ε. Note that we restrict the memory parameters of both X_{it} and ε_{it} to be identical across individuals i, so that the cross-sectional heterogeneity in memory is replaced by the cross-sectionally heterogeneous effect of long memory factors induced by factor loadings. This sacrifices generality to some extent but makes it more convenient to handle the model in theory.

There are different ways to define a long memory process in the literature (see Haldrup and Vald'es, 2017), among which a widespread one is to define it by a linear process (see e.g. Robinson and Hidalgo, 1997). To be specific, when d_ε and all of the elements of d_F and d_X are within $[0, \frac{1}{2})$, then F_t, X_{it} and ε_{it} have the following moving average representation:

$$F_t = \mu_F + \sum_{j=0}^{\infty} A_{F,j}\zeta_{F,t-j} \equiv \mu_F + F_t^o, \tag{3.2.2}$$

$$X_{it} = \mu_{X,i} + \sum_{j=0}^{\infty} A_{X,j}\zeta_{X,i,t-j} \equiv \mu_{X,i} + X_{it}^o \text{ for } i = 1, \ldots, N, \tag{3.2.3}$$

$$\varepsilon_{it} = \sum_{j=0}^{\infty} A_{\varepsilon,j}\zeta_{\varepsilon,i,t-j}, \tag{3.2.4}$$

where $A_{F,j}$ and $A_{X,j}$ are respectively $R \times R$ and $P \times P$ coefficient matrices and $A_{\varepsilon,j}$ is a scalar, and $\zeta_{F,t}$, $\zeta_{X,i,t}$ and $\zeta_{\varepsilon,i,t}$ are the corresponding innovation processes; and μ_F and $\mu_{X,i}$ are respectively $R \times 1$ and $P \times 1$ vectors of expectation. This specification of long memory processes includes the stationary $ARFIMA\,(p, d, q)$ as a special case. And different from $\zeta_{F,t-j}$, the innovations of regressors and idiosyncratic error term characterize the heterogeneity and dependence of X_{it} and ε_{it} across both individuals and time periods; see Section 3.3 for more detail. Following Bai (2009), the least squares (LS hereafter) estimator of β and F_t in time domain are given by the solution of the following nonlinear equations:

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i'\mathbf{M}_{\hat{F}}X_i\right)^{-1} \sum_{i=1}^{N} X_i'\mathbf{M}_{\hat{F}}Y_i \tag{3.2.5}$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\hat{\beta}\right)\left(Y_i - X_i\hat{\beta}\right)'\right]\hat{F} = \hat{F}V_{NT}, \tag{3.2.6}$$

together with restrictions $\frac{F'F}{T} = \mathbb{I}_R$ and $\frac{\Lambda'\Lambda}{N}$ being a diagonal matrix, where $F = (F_1', \ldots, F_T')'$ and $\Lambda = (\lambda_1', \ldots, \lambda_N')'$. In this study, we mainly focus on the asymptotic behavior of $\hat{\beta}$, which is discussed in the next section.

3.3 Asymptotic Behavior of Least Squares Estimator

In this section, we analyze the asymptotic behavior of the LS estimator in time domain given by (3.2.5) and (3.2.6). In the following we will use β^0, F_t^0, and F^0 to denote the true values of β, F_t, and F, respectively. We continue to use λ_i to denote the true value of the factor loadings as it is not directly estimated or involved in our theoretical analysis.

To proceed, we further introduce some notation. Define

$$D_{NT}(F) = \frac{1}{NT}\sum_{i=1}^{N} X_i'\mathbf{M}_F X_i - \frac{1}{T}\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i'\mathbf{M}_F X_k a_{ik}\right] \equiv \frac{1}{NT}\sum_{i=1}^{N} Z_i(F)' Z_i(F),$$

where $a_{ik} = \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$, and $Z_i(F) = \mathbf{M}_F X_i - \frac{1}{N}\sum_{k=1}^{N} a_{ik}\mathbf{M}_F X_k = (Z_{i1}(F), ..., Z_{iT}(F))'$. This matrix is important in the asymptotic representation of $\hat{\beta} - \beta^0$ and is also defined in Bai (2009, pp. 1240). Let $Z_i = Z_i(F^0) = (Z_{i1}, ..., Z_{iT})'$ and $D_{NT} = D_{NT}(F^0)$. Also for innovations let $\zeta_{X,t} = (\zeta_{X,1,t}, \ldots, \zeta_{X,N,t})'$ and $\zeta_{\varepsilon,t} = (\zeta_{\varepsilon,1,t}, \ldots, \zeta_{\varepsilon,N,t})'$. And let $\gamma_N(s,t) = \frac{1}{N}\sum_{i=1}^{N} E(\varepsilon_{it}\varepsilon_{is})$. For maximal memory parameters let $d_{X,\max} = \max_{1\le p\le P} d_{X_p}$ and $d_{F,\max} = \max_{1\le r\le R} d_{F_r}$. Similarly let $d_Z = (d_{Z_1}, \ldots, d_{Z_P})'$ be the memory parameter of Z_{it}, and $d_{Z,\max} = \max_{1\le p\le P} d_{Z_p}$. Then for minimal memory parameters let $d_{X,\min} = \min_{1\le p\le P} d_{X_p}$, and similarly we can define $d_{F,\min}$ and $d_{Z,\min}$. Let M be a generic positive constant that may vary across places.

In the following we introduce some technical assumptions together with remarks about them.

Assumption A. (i) When each element of $d_a \in (0, \frac{1}{2})$, then $A_{a,j} \sim Diag\left(j^{d_a-1}\right)\Pi_a$ as $j \to \infty$ for $a = F, X, \varepsilon$, where $Diag\left(j^{d_a-1}\right)$ is a diagonal matrix (or scalar if $a = \varepsilon$) with the main diagonal elements given by $j^{d_{F_r}-1}$ for $r = 1, \ldots, R$, or $j^{d_{X_p}-1}$ for $p = 1, \ldots, P$, or $j^{d_\varepsilon-1}$; and the $R \times R$ matrix Π_F and $P \times P$ matrix Π_X and scalar Π_ε that are all nonsingular. Otherwise we assume $A_{a,j}$ is square summable in certain norm.

(ii) $\zeta_{F,t}$, $\zeta_{X,t}$ and $\zeta_{\varepsilon,t}$ satisfy $E\left(\zeta_{F,t}|\mathcal{F}_{F,t-1}\right) = 0$, $E\left(\zeta_{X,t}|\mathcal{F}_{X,t-1}\right) = 0$, and $E\left(\zeta_{\varepsilon,t}|\mathcal{F}_{\varepsilon,t-1}\right) = 0$, where $\mathcal{F}_{F,t-1}$, $\mathcal{F}_{X,t-1}$ and $\mathcal{F}_{\varepsilon,t-1}$ are the corresponding filtrations.

(iii) Let $\zeta_{F,t(p)}$ be the p-th element of $\zeta_{F,t}$, and the same notation applies to $\zeta_{X,t}$. We assume that $\zeta_{F,t}$ satisfy

$$E\left[\zeta_{F,t(p)}\zeta_{F,t(q)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{1,pq} < \infty,$$

$$E\left[\zeta_{F,t(p_1)}\zeta_{F,t(p_2)}\zeta_{F,t(p_3)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{2,p_1p_2p_3} < \infty,$$

and

$$E\left[\zeta_{F,t(p_1)}\zeta_{F,t(p_2)}\zeta_{F,t(p_3)}\zeta_{F,t(p_4)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{3,p_1\ldots p_4} < \infty$$

for some absolute constants $\Phi_{1,pq}$, $\Phi_{2,p_1p_2p_3}$ and $\Phi_{3,p_1\ldots p_4}$, and for arbitrary p-, q- and p_1-, $\ldots$, p_4-th elements of $\zeta_{F,t}$. Also the same condition holds for $\zeta_{X,t}$ and $\zeta_{\varepsilon,t}$. Additionally, $\zeta_{\varepsilon,t}$ satisfies the following eighth-order moment condition

$$E\left[\zeta_{\varepsilon,t(p_1)}\cdots\zeta_{\varepsilon,t(p_8)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{4,p_1\ldots p_8} < \infty \tag{3.3.1}$$

for some absolute constant $\Phi_{4,p_1\ldots p_8}$, and for arbitrary p_1-, $\ldots$, p_8-th element of $\zeta_{\varepsilon,t}$.

(iv) $\zeta_{\varepsilon,i,t}$ is independent of $\zeta_{X,i,s}$, $\zeta_{F,r}$ and λ_j for all $r, s, t = 1, \ldots, T$ and $i, j = 1, \ldots, N$.

Assumption B. (i) $E\left\|X_{it}\right\|^4 \leq M$.

(ii) Let $\mathcal{F} = \left\{F \in \mathbb{R}^{T\times R} : F'F/T = \mathbb{I}_R\right\}$. We assume $\inf_{F\in\mathcal{F}} D_{NT}(F) > 0$.

(iii) $E\left\|F_t^0\right\|^4 \leq M$ and $\frac{1}{T}F^{0\prime}F^0 \xrightarrow{p} \Sigma_F > 0$ for some $R \times R$ matrix Σ_F, as $T \to \infty$.

(iv) $E\left\|\lambda_i\right\|^4 \leq M$ and $\frac{1}{N}\Lambda'\Lambda \xrightarrow{p} \Sigma_\Lambda > 0$ for some $R \times R$ matrix Σ_Λ, as $N \to \infty$.

Assumption C. (i) $E(\varepsilon_{it}) = 0$ and $E\left|\varepsilon_{it}\right|^8 \leq M$.

(ii) $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij,ts}$, $\left|\sigma_{ij,ts}\right| \leq \overline{\sigma}_{ij}$ for all (t, s), $\left|\sigma_{ij,ts}\right| \leq \tau_{ts}$ for all (i, j),

$$\frac{1}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij} \leq M, \tag{3.3.2}$$

and

$$\frac{1}{T^{1+2d_\varepsilon}}\sum_{t,s=1}^{T}\tau_{ts} \leq M,\ \frac{1}{NT^{1+2d_\varepsilon}}\sum_{i,j,t,s=1}\left|\sigma_{ij,ts}\right| \leq M,\ \frac{1}{T^{\max(4d_\varepsilon,1)}}\sum_{t,s=1}^{T}\left|\gamma_N(s,t)\right|^2 \leq M. \tag{3.3.3}$$

(iii) For every (t, s), $E\left|N^{-\frac{1}{2}}\sum_{i=1}^{N}\left[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})\right]\right|^4 \leq M$.

(iv) Moreover

$$\frac{1}{NT^{1+2d_\varepsilon}}\sum_{i,k=1}^{N}\sum_{t,s=1}^{T}\left|\operatorname{cov}(\varepsilon_{it}\varepsilon_{is}, \varepsilon_{kt}\varepsilon_{ks})\right| \leq M,$$

$$\frac{1}{N^2T^{1+2d_\varepsilon}}\sum_{t,s=1}^{T}\sum_{i,j,k,l=1}^{N}\left|\operatorname{cov}(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{ls})\right| \leq M,$$

and

$$\frac{1}{NT^{2+4d_\varepsilon}}\sum_{i,k=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\operatorname{cov}(\varepsilon_{it}\varepsilon_{is}, \varepsilon_{ku}\varepsilon_{kv})\right| \leq M.$$

Assumption D. (i) Suppose $E(F_t^o F_s^{o\prime}) = \Sigma_{F,ts}$ and $\left\|\Sigma_{F,ts}\right\| \leq \tau_{F,ts}$. We assume $\frac{1}{T^{1+2d_{F,\max}}}\sum_{t,s=1}^{T}\tau_{F,ts} \leq M$, and

$$\frac{1}{T^{\max(2d_\varepsilon+2d_{F,\max},1)}}\sum_{t,s=1}^{T}\tau_{ts}\tau_{F,ts} \leq M \text{ and } \frac{1}{NT^{\max(2d_{F,\max}+2d_\varepsilon,1)}}\sum_{i,j,t,s=1}\left|\sigma_{ij,ts}\right|\tau_{F,ts} \leq M. \tag{3.3.4}$$

(ii) Suppose $E(\chi_{it}^{o}\chi_{js}^{o\prime}) = \Sigma_{X,ijts}$ with $\chi = X, Z$; and $\|\Sigma_{\chi,ijts}\| \leq \tau_{\chi,ts}$ for all (i,j). With $\sigma_{\chi,ijts} = \mathrm{tr}\left(\Sigma_{\chi,ijts}\right)$, we assume $\frac{1}{T^{1+2d_{\chi,\max}}}\sum_{t,s=1}^{T}\tau_{\chi,ts} \leq M$, and

$$\frac{1}{T^{\max(2d_{\chi,\max}+2d_{\varepsilon},1)}}\sum_{t,s=1}^{T}\tau_{ts}\tau_{\chi,ts} \leq M \text{ and } \frac{1}{N^2T^{\max(2d_{\chi,\max}+2d_{\varepsilon},1)}}\sum_{i,j,k,l=1}^{N}\sum_{t,s=1}^{T}|\sigma_{ij,ts}|\,|\sigma_{\chi,ijts}| \leq M.$$

Assumption E. (i) $\mathrm{plim}\, D\left(F^0\right) = D_0$ for some nonrandom positive definite matrix D_0.

(ii) Suppose F^0 does not contain any constant column, $N^{-\frac{1}{2}}T^{d_{\varepsilon}-\frac{1}{2}}\sum_{i=1}^{N}Z_i'\varepsilon_i \xrightarrow{d} \mathcal{N}\left(0,\Sigma\right)$;

(iii) Suppose there exists constant column in F^0, then

$$N^{-\frac{1}{2}}T^{\max(d_{Z,\max}+d_{\varepsilon},1/2)-1}\sum_{i=1}^{N}Z_i'\varepsilon_i \xrightarrow{d} \mathcal{N}\left(0,\Sigma\right),$$

where

$$\mathrm{p}\lim \rho_{NT}^{-2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\sigma_{ij,ts}Z_{it}Z_{js}' = \Sigma, \text{ and } \rho_{NT} = N^{\frac{1}{2}}T^{\frac{1}{2}-d_{\varepsilon}}$$

in case (ii) and

$$\rho_{NT} = N^{\frac{1}{2}}T^{1-\max(d_{Z,\max}+d_{\varepsilon},1/2)}$$

in case (iii).

Assumption F. $d_{\varepsilon} \leq \min\left\{d_{F,\min}, d_{X,\min}\right\}$.

Remark 1. Assumption A is a panel data extension to the classic setting of a stationary long memory linear process (see e.g. Nielsen, 2005). To be specific, the first half of Assumption A(i) is adopted from Chung (2002), whose Lemma 2 shows that autocovariance of F_t^0, X_{it} and ε_{it} satisfy, as $j \to \infty$

$$\Gamma_F\left(j\right) = \mathrm{Cov}\left(F_t^0, F_{t-j}^0\right) \sim Diag\left(j^{d_F-\frac{1}{2}}\right)C_F Diag\left(j^{d_F-\frac{1}{2}}\right),$$

$$\Gamma_{X_i}\left(j\right) = \mathrm{Cov}\left(X_{it}, X_{i,t-j}\right) \sim Diag\left(j^{d_X-\frac{1}{2}}\right)C_X Diag\left(j^{d_X-\frac{1}{2}}\right),$$

and

$$\Gamma_{\varepsilon_i}(j) = Cov(\varepsilon_{it}, \varepsilon_{i,t-j}) \sim C_\varepsilon j^{2d_\varepsilon - 1}, \tag{3.3.5}$$

for some absolute constant matrices C_F, C_X and scalar C_ε. The above approximations imply the square summability of $\|A_{a,j}\|$ for $a = F^0$, X and ε. Take $A_{F^0,j}$ for instance. Assumption A(i) implies that for any $\delta > 0$, there exists an integer $K_\delta > 0$ such that $\|A_{F^0,j}\|^2 \leq (1+\delta)^2 C \sum_{r=1}^{R} j^{2d_{F_r}-2}$ for some positive constant C when $j \geq K_\delta$, which then implies

$$\sum_{j=0}^{\infty} \|A_{F^0,j}\|^2 = \sum_{j=0}^{K_\delta} \|A_{F^0,j}\|^2 + \sum_{j=K_\delta}^{\infty} \|A_{F^0,j}\|^2 \leq C_\delta + (1+\delta)^2 C \sum_{r=1}^{R} \sum_{j=K_\delta}^{\infty} j^{2d_{F_r}-2} < \infty,$$

by Riemann sum approximation if $d_{F_r} < \frac{1}{2}$. This illustrates how A(i) defines a stationary long memory process through the hyperbolic rate of decay of its autocovariance function. Note that this part of Assumption A(i) only covers the long memory scenario as it emphasizes the hyperbolic rate of decay of autocovariance function, while for short memory processes like ARMA model, the rate is usually exponential and thus not nested in this half of Assumption A(i) by simply substituting $d_a = 0$. Therefore in the second half we assume the stationarity of all the variables when some short memory processes are involved. We can see in Assumption C about how we uniformly deal with short and long memory, which is explained in Remark 3. As we mentioned before, a widespread alternative definition of long memory process is modeling it by a fractionally integrated process $I(d)$, which can be extended to an $ARFIMA(p,d,q)$ model that is popular in application. Relative to the fully parametric definition of $ARFIMA(p,d,q)$, ours is termed as semiparametric modeling of long memory processes, which is free from short-run dynamics specification and thus can avoid inconsistent estimation if we misspecify the model, such as the autoregressive or moving average parts. In Assumption A(ii) and A(iii), we impose moment conditions up to the eighth order. And Assumption A(iv) implies that ε_{it} is independent of X_{js}, λ_j, and F_s^0 for all i,

t, j, and s, which is also assumed in Bai (2009).

Remark 2. Assumption B is also borrowed from Bai (2009), specifying a finite fourth-order moment for both the factors and factor loadings and a restriction of strong factors. Note that both of the moment conditions in Assumption B(i) and B(iii) can be justified by the corresponding fourth-order moment conditions in Assumption A(i) of innovations.

Remark 3. Assumption C(i) and C(ii) can be implied by our Assumption A. The reason why we separately list these two sets of assumptions is that our Assumption A is comparable to the standard definition of stationary long memory process, while our Assumption C is comparable to the corresponding Assumption C in Bai (2009). To be specific, C(i) is implied by Assumption A(ii) and A(iii). To see this, Assumption A(ii) implies the zero expectation and (3.3.1) in Assumption A(iii) together with the square summability indicated by A(i) can imply the finite eighth-order moment. In Assumption C(ii), (3.3.2) is the standard condition of cross-sectional weak dependence of $\varepsilon_{i,t}$. And the other inequalities specify the serial dependence, as they generalize the Assumption C(ii) in Bai (2009) by including long memory. The idea is adopted from the Theorem 1 of Chung (2002) via a direct application of (3.3.5). To see this, we consider the bound $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i,j). Consider the simplest case where $i=j$, we have $\sigma_{ii,ts} = \sigma_{ii,t-s}$ by its stationarity, and we can express the bound $\tau_{ts} = \tau_{t-s}$ accordingly. By symmetry of τ_{t-s} as $\tau_{t-s} = \tau_{s-t}$,

$$\frac{1}{T}\sum_{t,s=1}^{T}\tau_{ts} = \frac{1}{T}\sum_{t,s=1}^{T}\tau_{t-s} = \tau_0 + \frac{2}{T}\sum_{t>s}\tau_{t-s} = \frac{2}{T}\sum_{k=1}^{T-1}(T-k)\,\tau_k + O(1)\,. \quad (3.3.6)$$

Let $\gamma_i(k)$ be an arbitrary autocovariance function of order k of ε_{it}. By (3.3.5) $\gamma_i(k) \sim C_\varepsilon k^{2d_\varepsilon - 1}$ for some constant C_ε as $k \to \infty$. Then for any $\delta > 0$, there exists an integer $K_\delta > 0$ such that $(1-\delta)\,C_\varepsilon k^{2d_\varepsilon-1} \leq \gamma_i(k) \leq (1+\delta)\,C_\varepsilon k^{2d_\varepsilon-1}$ when $k \geq K_\delta$. Let $\tau_k = |\gamma(k)|$ be an appropriate upper bound for $|\gamma_i(k)|$

uniformly over $i = 1, \ldots, N$. We have

$$\begin{aligned}
\frac{1}{T}\sum_{k=1}^{T-1}(T-k)\tau_k &= \sum_{k=1}^{K_\delta}\left(1-\frac{k}{T}\right)|\gamma(k)| + \sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)|\gamma(k)| \\
&= \sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)|\gamma(k)| + O(1) \\
&\geq C_\varepsilon(1-\delta)\sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)k^{2d_\varepsilon-1} + O(1) \\
&= C_\varepsilon(1-\delta)T^{2d_\varepsilon}\frac{1}{T}\sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)\left(\frac{k}{T}\right)^{2d_\varepsilon-1} + O(1) \\
&= C_\varepsilon(1-\delta)T^{2d_\varepsilon}\int_{K_\delta/T}^{1}(1-r)\,r^{2d_\varepsilon-1}dr\left\{1+O\left(\frac{1}{T}\right)\right\} + O(1),
\end{aligned} \tag{3.3.7}$$

given the convergence of both $\int_{K_\delta/T}^{1}(1-r)\,r^{2d_\varepsilon-1}dr$ and $\sum_{k=1}^{K_\delta}(1-\frac{k}{T})\,|\gamma(k)|$ when $d_\varepsilon > 0$. The above calculations indicate that the condition $\frac{1}{T}\sum_{t,s=1}^{T}\tau_{ts} \leq M$ in Bai (2009) is generally violated unless $d_\varepsilon = 0$. The same reasoning applies to show the second inequality in (3.3.3) as long as the cross-sectional correlations among $\{\varepsilon_{it}\}$ are "weak enough". Analogously, for the third inequality in (3.3.3), we have

$$\begin{aligned}
\frac{1}{T}\sum_{k=1}^{T-1}(T-k)\tau_k^2 &\approx C_\varepsilon T^{4d_\varepsilon-1}\frac{1}{T}\sum_{k=1}^{T-1}\left(1-\frac{k}{T}\right)\left(\frac{k}{T}\right)^{4d_\varepsilon-2} \\
&= C_\varepsilon T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}(1-r)\,r^{4d_\varepsilon-2}dr\left\{1+O\left(\frac{1}{T}\right)\right\} + O(1),
\end{aligned}$$

given the convergence of the last integral, which requires $d_\varepsilon > 1/4$ so that $4d_\varepsilon - 2 > -1$. When $0 \leq d_\varepsilon \leq 1/4$, we notice that

$$T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}(1-r)\,r^{4d_\varepsilon-2}dr = T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}r^{4d_\varepsilon-2}dr - T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}r^{4d_\varepsilon-1}dr, \tag{3.3.8}$$

where the second integral is convergent. And the first integral is further given by

$$T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1} r^{4d_\varepsilon-2}dr = \int_{K_\delta/T}^{1}(rT)^{4d_\varepsilon-2}\,d(rT) \equiv \int_{K_\delta}^{T}(r_*)^{4d_\varepsilon-2}\,d(r_*) \quad (3.3.9)$$

$$= T^{4d_\varepsilon-1} + O(1) = O(1). \quad (3.3.10)$$

It follows that $\frac{1}{T^{\max(1,4d_\varepsilon)}}\sum_{k=1}^{T-1}(T-k)\,\tau_k^2 \leq M$, which implies the last condition in (3.3.3). In the special case where $d_\varepsilon = 0$, Assumption C(ii) degenerates to Assumption C(ii) in Bai (2009) which involves only short-range dependence. Although in this case the integral derived in the end of (3.3.7) is not convergent, the moment condition still coincides with the one in Bai (2009). One special case of our setup is a linear process with fractional integration, like $(1-L)^{d_F}\,F_t^0 = e_t$ with L the lag-operator and e_t a short memory process. Assumption C(iii), which reflects the weak cross-sectional dependence, is directly borrowed from the Assumption C(iii) in Bai (2009). With more tedious arguments, one can also verify Assumption C(iv), as it extends the higher-order moment conditions of short memory process in the Assumption C(iv) in Bai (2009). We omit them here for brevity.

Remark 4. Assumption D(i) and D(ii) are adopted from the convergence rate indicated by Theorem 3 in Chung (2002), where by construction of Z_{it} it could be treated as a potentially long memory process as well. To provide an intuitive explanation of the whole Assumption D, we take F_t^0 for instance. Assume $R=1$ and let $\gamma_F(k)$ denote the autocovariance of F_t^0. Then $\gamma_F(k) \sim C_F k^{2d_F-1}$ for some constant C_F as $k\to\infty$. Then following the reasoning in (3.3.6) and (3.3.7), we have

$$\frac{1}{2T}\sum_{t,s=1}^{T}\tau_{ts}\tau_{F,ts}$$

$$= \frac{1}{T}\sum_{k=1}^{T-1}(T-k)\,\tau_k\tau_{F,k} + O(1)$$

$$
\begin{aligned}
&= \sum_{k=1}^{K_\delta}\left(1-\frac{k}{T}\right)|\gamma(k)||\gamma_F(k)| + \sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)|\gamma(k)||\gamma_F(k)| + O(1) \\
&= \sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)|\gamma(k)||\gamma_F(k)| + O(1) \\
&\geq C_\varepsilon C_F(1-\delta)^2 \sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)k^{2d_\varepsilon-1}k^{2d_F-1} + O(1) \\
&= C_\varepsilon C_F(1-\delta)^2 T^{2d_\varepsilon+2d_F-1}\frac{1}{T}\sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)\left(\frac{k}{T}\right)^{2d_\varepsilon+2d_F-2} + O(1) \\
&= C_\varepsilon C_F(1-\delta)^2 T^{2d_\varepsilon+2d_F-1}\int_{K_\delta/T}^{1}(1-r)\,r^{2d_\varepsilon+2d_F-2}dr\left\{1+O\left(\frac{1}{T}\right)\right\} + O(1).
\end{aligned}
$$

The integral in the last equality is convergent only when $d_\varepsilon+d_F > \frac{1}{2}$. When $d_\varepsilon+d_F \leq 1/2$, we can readily show $T^{2d_\varepsilon+2d_F-1}\int_{K_\delta/T}^{1}(1-r)\,r^{2d_\varepsilon+2d_F-2}dr = O(1)$ by the same reasoning in (3.3.8)-(3.3.10). It follows that $\frac{1}{T^{\max(2d_\varepsilon+2d_F,1)}}\sum_{t,s=1}^{T}\tau_{ts}\tau_{F,ts} \leq M$ and the first part of (3.3.4) in Assumption D(i) holds. Similarly, the second part of (3.3.4) also holds provided the cross-sectional correlations are sufficiently weak.

Remark 5. Assumption E(i) corresponds partly to Assumption E in Bai (2009), giving probabilistic limit of $D(F)$ required for asymptotic covariance matrix of $\hat{\beta}-\beta^0$. Assumption E(ii) gives the convergence rate of $\hat{\beta}-\beta^0$. Because cross-sectional weak dependence of ε_{it} indicates we can use Lindeberg-Lévy CLT over i, which requires a uniform boundedness of the second moment $E(Z_i'\varepsilon_i\varepsilon_i'Z_i) = \sum_{t,s=1}^{T}E(\varepsilon_{it}\varepsilon_{is})E(Z_{it}Z_{is}')$ by definition of Z_i and Assumption A(iv). As we will see in the following, whether $E(Z_i)=0$ for all i is determinant in convergence rate. Together with data generating processes in (3.2.2)-(3.2.4) and the strict exogeneity condition in Assumption A(iv), we can see how temporal dependence of $Z_{it}\varepsilon_{it}$ is dominated by the mean of Z_{it}, denoted by μ_Z when they are nonzero. To illustrate this idea in a simple manner, consider $\sum_{t=1}^{T}Z_{it}\varepsilon_{it}$ for arbitrary i. Its mean is zero and its variance-covariance matrix is given by

$$E\left[\left(\sum_{t=1}^{T} Z_{it}\varepsilon_{it}\right)\left(\sum_{t=1}^{T} Z_{it}\varepsilon_{it}\right)'\right] = \sum_{t,s=1}^{T} E\left(Z_{it}Z_{is}'\right)E\left(\varepsilon_{it}\varepsilon_{is}\right)$$
$$= \sum_{t,s=1}^{T} \left[\mu_Z\mu_Z' + E\left(Z_{it}^{o}Z_{is}^{o\prime}\right)\right]E\left(\varepsilon_{it}\varepsilon_{is}\right),$$

where Z_{it}^{o} is defined in the same way as X_{it}^{o} in (3.2.4). By Assumption C(ii) and D(ii),

$$\sum_{t,s=1}^{T} \mu_Z\mu_Z' E\left(\varepsilon_{it}\varepsilon_{is}\right) = O\left(T^{1+2d_\varepsilon}\right) \text{ and } \sum_{t,s=1}^{T} E\left(Z_{it}^{o}Z_{is}^{o\prime}\right)E\left(\varepsilon_{it}\varepsilon_{is}\right) = O\left(T^{\max(2d_{Z,\max}+2d_\varepsilon,1)}\right),$$

thus $\mu_F\mu_F'$ dominates in the summation as along as $\mu_F \neq 0$, and only the autocovariance structure of ε_{it} is applicable because of its mean-zero nature. If $\mu_F = 0$, the order above will be also affected by $d_{F,\max}$ as Assumption D(i) implies. Note that by definition, Z_i can be interpreted as the residual of linear projection of X_i on the column space of F, and demeaned by a weighted average. So by construction $E(Z_i'F^0) = 0$ holds by orthogonality, and $E\left(Z_i\right) = 0$ if F^0 contains a constant column or if

$$X_{it} = \phi_i F_t^0 + u_{it}$$

with $E\left(u_{it} \mid F^0\right) = 0$ is the true data generating process, that is to say, (X_{it}, F_t^0) follow a linear regression model or X_{it} follows a pure factor model with latent factors given by F_t. The latter setting is adopted in some current studies (Ergemen, 2019, among others) and is more restrictive but easier to deal with in practice, so in this study we only focus on the former setting that F contains a constant column. If $E\left(Z_i\right) = 0$, the convergence rate is adopted from Theorem 3 in Chung (2002). In pure time series models, we cannot obtain asymptotic normality for OLS estimator when $d_{Z,\max} + d_\varepsilon \geq \frac{1}{2}$, but in panel models, weak dependence over i and large number of cross-sectional units can help us regain the asymptotic normality. Moreover, a constant column in F^0 indicates the existence of individual fixed effect in our model. In Bai (2009) this is only a special case of factor as LS estimator is only less efficient when individual fixed effect is not canceled out by within-group transformation. But in our model, individual fixed effect may affect the convergence rate of LS estimator when long memory exists in idiosyncratic error term. One may think about dealing with such issue by doing within-group transformation in our model before the LS estimation. In our online supplement we will discuss how the asymptotic behavior of $\hat{\beta}$ changes under such transformation.

Remark 6. Assumption F indicates the condition of fractional cointegration, which generalizes the usual cointegration notion in time series literature, see e.g. Marinucci and Robinson (2001). It also implies that $d_\varepsilon < d_{Z,\min}$ by construction of Z_i.

Let F_{rt}^0 and $Z_{k,it}$ denote the r-th and k-th element of F_t^0 and Z_{it}. The following theorem establishes the asymptotic distribution of the LS estimator $\hat{\beta}$.

Theorem 3.3.1 *Suppose that Assumption A-F hold. Then for comparable N and T such that $T/N \to \rho > 0$, we have*

$$\rho_{NT}\left(\hat{\beta} - \beta^0 - \frac{1}{T^{1-2d_\varepsilon}}A_{NT} - \frac{1}{N}C_{NT}\right) \xrightarrow{d} \mathcal{N}\left(0, D_0^{-1}\Sigma D_0^{-1}\right),$$

where ρ_{NT} is defined in Assumption E(ii) that depends on the setup for the factor F^0 and magnitude of $d_{Z,\max} + d_\varepsilon$, D_0 and Σ are given in Assumption E(i), and the bias terms A_{NT} and C_{NT} are each $O_p(1)$ and given by

$$A_{NT} = -D_{NT}^{-1}\frac{1}{NT^{1+2d_\varepsilon}}\sum_{i=1}^{N} X_i'\mathbf{M}_{F^0}\frac{1}{N}\sum_{k=1}^{N}\Omega_k\hat{F}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,$$

and

$$C_{NT} = -D_{NT}^{-1} \frac{1}{N} \sum_{i=1}^{N} \frac{(X_i - V_i)' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \frac{1}{T} \sum_{k=1}^{N} \lambda_k \varepsilon_k' \varepsilon_i,$$

where $V_i = \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k$ *with* $a_{ik} = \lambda_i' \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \lambda_k$.

The above theorem shows in factor model, the convergence rate of the LS estimator $\hat{\beta}$ is slowed by the existence of long memory. In terms of limiting distribution, although we still have asymptotic normality, the bias terms now have their orders dependent on the long memory parameters. In the special case where all memory parameters are zero, the above result is the same as the one obtained by Bai (2009), which shows how our Theorem 3.3.1 nests the short memory setting as a special case. However, the convergence rate ρ_{NT} has a complex representation based on whether F^0 has a constant column, and whether $d_{Z,\max} + d_\varepsilon$ is greater than $\frac{1}{2}$ or not. This dramatically complicates the implementation of LS estimator, which is illustrated by Monte Carlo simulations in Section 3.6, where we find out the performance of analytical bias correction is poor. In the meantime we adopt the half-panel jackknife bias correction that is adjusted by memory parameters, whose results are not good either after we plug in the local Whittle estimator of the memory parameters. This difficulty in implementation calls for an alternative method to deal with stationary long memory in our model. In the next section we try to develop a frequency domain least squares estimator that is widely studied in long memory time series regression model, and analyze its asymptotic behavior in our panel setup.

3.4 Frequency Domain Least Squares Estimator

In this section we introduce the frequency domain least squares (FDLS) estimator and then study its asymptotic properties.

3.4.1 Estimation method

From Theorem 3.3.1, we can see how complicated the implementation is for LS estimator to account for the potential existence of long memory. To simplifies the implementation, we propose a new estimator, which is based on the LS estimation over frequency domain (FDLS hereafter). To be specific, we consider the discrete Fourier transform (DFT hereafter) on both sides of model (3.2.1) over frequency γ_j,

$$\begin{aligned}&\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}Y_{it}e^{\mathbf{i}t\gamma_j}\\&=\frac{\beta'}{\sqrt{2\pi T}}\sum_{t=1}^{T}X_{it}e^{\mathbf{i}t\gamma_j}+\frac{1}{\sqrt{2\pi T}}\lambda_i'\sum_{t=1}^{T}F_t^0e^{\mathbf{i}t\gamma_j}\\&+\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\varepsilon_{it}e^{\mathbf{i}t\gamma_j},\ i=1,\ldots,N,\ j=1,\ldots,L.\end{aligned}\tag{3.4.1}$$

where $\mathbf{i}=\sqrt{-1}$ is the imaginary unit, and $\gamma_j=\frac{2\pi j}{T}$ for $j=1,\ldots,L$. The frequencies γ_j are called "Fourier frequencies", which removes the mean of the processes in the frequency domain. For instance, consider the DFT over Fourier frequencies of F_t^0 given by (3.2.2), we have

$$\begin{aligned}\sum_{t=1}^{T}F_t^0e^{\mathbf{i}t\gamma_j}&=\mu_F\sum_{t=1}^{T}e^{\mathbf{i}t\gamma_j}+\sum_{t=1}^{T}F_t^oe^{\mathbf{i}t\gamma_j}\\&=\mu_F\frac{e^{\mathbf{i}\gamma_j}\left(1-e^{\mathbf{i}T\gamma_j}\right)}{1-e^{\mathbf{i}\gamma_j}}+\sum_{t=1}^{T}F_t^oe^{\mathbf{i}t\gamma_j}=\sum_{t=1}^{T}F_t^oe^{\mathbf{i}t\gamma_j}\end{aligned}$$

by Euler's identity $e^{\mathbf{i}T\gamma_j}=e^{\mathbf{i}2\pi j}=1$.

For ease of notation, let $W_{Y,ij} = \sum_{t=1}^{T} Y_{it} e^{\mathrm{i}t\gamma_j}$, and define $W_{X,ij}$, $W_{F,j}$ and $W_{\varepsilon,ij}$ analogously. Let $W_{a,i} = \left(W'_{a,i1}, \ldots, W'_{a,iL}\right)'$ for $a = Y, X, \varepsilon$ and $W_F = \left(W'_{F,1}, \ldots, W'_{F,L}\right)'$. Note that $W_{Y,i}$, $W_{X,i}$, and W_F are $L \times 1$, $L \times P$, and $L \times R$ matrices, respectively. Then (3.4.1) can be rewritten as

$$W_{Y,ij} = \beta' W_{X,ij} + \lambda_i' W_{F^0,j} + W_{\varepsilon,ij},\ i = 1, \ldots, N,\ j = 1, \ldots, L. \tag{3.4.2}$$

This model can also be treated as a panel data model, with T time periods replaced by L frequencies. Then the FDLS estimator has the following objective function

$$\begin{aligned} SSR\left(\beta, W_F, \Lambda\right) &= \frac{1}{NT} \sum_{i=1}^{N} \left(W_{Y,i} - W_{X,i}\beta - W_F\lambda_i\right)^* \left(W_{Y,i} - W_{X,i}\beta - W_F\lambda_i\right) \\ &= \frac{1}{NT} \left\|W_Y - W_X \cdot \beta - \Lambda W_F'\right\|^2 \end{aligned} \tag{3.4.3}$$

subject to the constraint that $\tilde{\Gamma}_F W_F^* W_F \tilde{\Gamma}_F / T = \mathbb{I}_R$, where $\tilde{\Gamma}_F = Diag\left\{\gamma_L^{d_{F_r} - \frac{1}{2}}\right\}$. And $W_X \cdot \beta = \sum_{p=1}^{P} W_X^p \beta_p$ with β_p and W_X^p correspond to the p-th element of β and $W_{X,ij}$, and both W_Y and W_X^p are $N \times L$ complex matrices of DFT. Here $\tilde{\Gamma}_F$ is an $R \times R$ diagonal matrix for normalization over frequency domain, so as to make the notation consistent with the time domain setting. Such normalization can be justified by the properties of average periodogram, see our assumptions and remarks later on. We further denote $\tilde{W}_{F,j} = \tilde{\Gamma}_F W_{F,j}$ and $\tilde{\lambda}_i = \tilde{\Gamma}_F^{-1} \lambda_i$, by which we can rewrite the model (3.4.2) as

$$W_{Y,ij} = \beta' W_{X,ij} + \tilde{\lambda}_i' \tilde{W}_{F,j} + W_{\varepsilon,ij},\ i = 1, \ldots, N,\ j = 1, \ldots, L, \tag{3.4.4}$$

or in vector-matrix notation

$$W_{Y,i} = W_{X,i}\beta + \tilde{W}_F \tilde{\lambda}_i + W_{\varepsilon,i},\ i = 1, \ldots, N, \tag{3.4.5}$$

where $\tilde{W}_F = \left(\tilde{W}_{F,1}, \ldots, \tilde{W}_{F,L}\right)'$. Note that $\tilde{W}_F = W_F \tilde{\Gamma}_F$ and $\tilde{\Lambda} = (\tilde{\lambda}_1, ..., \tilde{\lambda}_N)' =$

$\Lambda\tilde{\Gamma}_F^{-1}$, and $\tilde{W}_F^*\tilde{W}_F/T = \mathbb{I}_R$ by construction. Define the projection matrix in a complex vector space by

$$\mathbf{M}_{\tilde{W}_F} = \mathbb{I}_L - \tilde{W}_F\left(\tilde{W}_F^*\tilde{W}_F\right)^{-1}\tilde{W}_F^* \equiv \mathbb{I}_L - \mathbf{P}_{\tilde{W}_F}.$$

Clearly, the columns of $\tilde{W}_F$ spans the same space as those of W_F because

$$\mathbf{P}_{\tilde{W}_F} = \tilde{W}_F\left(\tilde{W}_F^*\tilde{W}_F\right)^{-1}\tilde{W}_F^* = W_F\tilde{\Gamma}_F\tilde{\Gamma}_F^{-1}\left(W_F^*W_F\right)^{-1}\tilde{\Gamma}_F^{-1}\tilde{\Gamma}_F W_F^* = \mathbf{P}_{W_F}.$$

Then by construction $\mathbf{M}_{W_F}W_F = W_F^*\mathbf{M}_{W_F} = 0$. It follows that we can premultiply both sides of (3.4.5) by $\mathbf{M}_{\tilde{W}_F}$ to obtain

$$\mathbf{M}_{\tilde{W}_F}W_{Y,i} = \mathbf{M}_{\tilde{W}_F}W_{X,i}\beta + \mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i},\ i = 1,\ldots,N,$$

Then an infeasible FDLS estimator of β is obtained by regressing $\mathbf{M}_{\tilde{W}_F}W_{Y,i}$ on $\mathbf{M}_{\tilde{W}_F}W_{X,i}$ to obtain

$$\tilde{\beta}\left(\tilde{W}_F\right) = \left[\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\tilde{W}_F}W_{X,i}\right)\right]^{-1}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\tilde{W}_F}W_{Y,i}\right).$$

Next, we consider the infeasible FDLS estimation of the factors and factor loadings. Given β, we denote $U_i = U_i(\beta) = Y_i - X_i\beta$ and its DFT $W_{U,i}$ over the same Fourier frequencies as above. Then $W_{U,i}$ has the pure factor structure in frequency domain:

$$W_{U,i} = \tilde{W}_F\tilde{\lambda}_i + W_{\varepsilon,i}.$$

Define $W_U = (W_{U,1},\ldots,W_{U,N})'$ and $W_\varepsilon = (W_{\varepsilon,1},\ldots,W_{\varepsilon,N})'$, which are two $N \times L$ matrices. Then the FDLS objective function is

$$\frac{1}{NT}\mathrm{tr}\left[\left(W_U - \tilde{\Lambda}\tilde{W}_F'\right)^*\left(W_U - \tilde{\Lambda}\tilde{W}_F'\right)\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{L}\left|W_{U,ij} - \tilde{\lambda}_i'\tilde{W}_{F,j}\right|^2. \tag{3.4.6}$$

This objective function is identical to the one in Bai (2009, pp. 1236) except it is in the frequency domain, and we can concentrate $\tilde{\Lambda}$ out by using

$$\tilde{\Lambda} = W_U \overline{\tilde{W}}_F \left(\tilde{W}_F^* \tilde{W}_F\right)^{-1} = W_U \overline{\tilde{W}}_F / T \tag{3.4.7}$$

along with the restriction $\tilde{W}_F^* \tilde{W}_F / T = \mathbb{I}_R$. Using (3.4.7), the objective function in (3.4.6) becomes

$$\begin{aligned}
&\operatorname{tr}\left[\left(W_U - \tilde{\Lambda}\tilde{W}_F'\right)^* \left(W_U - \tilde{\Lambda}\tilde{W}_F'\right)\right] \\
&= \operatorname{tr}\left[\left(W_U - W_U \overline{\tilde{W}}_F \tilde{W}_F'/T\right)^* \left(W_U - W_U \overline{\tilde{W}}_F \tilde{W}_F'/T\right)\right] \\
&= \operatorname{tr}\left(W_U^* W_U\right) - \operatorname{tr}\left(\tilde{W}_F' W_U^* W_U \overline{\tilde{W}}_F\right)/T.
\end{aligned} \tag{3.4.8}$$

Therefore minimizing (3.4.6) is equivalent to maximizing $\operatorname{tr}\left(\tilde{W}_F' W_U^* W_U \overline{\tilde{W}}_F\right)$, which is the typical principal components analysis (PCA) problem in frequency domain, where $W_U^* W_U$ is the stacked periodogram of U. As documented in Brillinger (2001, pp. 70,342), PCA continues to work and the estimator of $\tilde{W}_F$, denoted by $\hat{W}_F$, is given by the eigenvectors multiplied by $\sqrt{T}$ of $W_U^* W_U$ corresponding to the its first R largest eigenvalues, which are real because $W_U^* W_U$ is Hermitian. Note that as in Bai (2009) the indeterminacy over rotation for $\tilde{W}_F$ still holds by the restriction $\tilde{W}_F^* \tilde{W}_F / T = \mathbb{I}_R$. Moreover, the above PCA actually obtains the estimator of $\tilde{W}_F$, which is normalized column-wise by the matrix $\tilde{\Gamma}_F$, so W_F is definitely not identified here, and the same issue holds for $\tilde{\Lambda}$. However, this lack of identifiability does not matter for our purpose in estimating β as we can see in the transformed model (3.4.4).

In practice, we iterate between β and $\tilde{W}_F$. So the feasible FDLS estimator $\left(\tilde{\beta}, \hat{W}_F\right)$ of $\left(\beta, \tilde{W}_F\right)$ is given by the solution of the following set of nonlinear equations:

$$\tilde{\beta} = \left[\sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{X,i}\right)\right]^{-1} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{Y,i}\right), \tag{3.4.9}$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)^{*}\right]\hat{W}_{F}=\hat{W}_{F}V_{NL}, \qquad (3.4.10)$$

where V_{NL} is the diagonal matrix containing the R largest eigenvalues of $W_U^* W_U$ in decreasing order.

In the next subsection we study the asymptotic properties of FDLS estimator.

3.4.2 Asymptotic properties of the frequency domain estimator

In this subsection, we develop the asymptotic theory for the FDLS estimator $\tilde{\beta}$ together with $\hat{W}_F$, which is the PCA estimator of DFT of the factor F. To proceed, we add some notation. Let $\tilde{\Gamma}_X = Diag\left\{\gamma_L^{d_{X_p}-\frac{1}{2}}\right\}$ and $\tilde{W}_{X,i} = W_{X,i}\tilde{\Gamma}_X$ for each i in the same manner as we define $\tilde{\Gamma}_F$ and $\tilde{W}_F$ above. Similarly, let $\tilde{\Gamma}_\varepsilon = \gamma_L^{d_\varepsilon-\frac{1}{2}}$ and $\tilde{W}_{\varepsilon_i} = W_{\varepsilon,i}\tilde{\Gamma}_\varepsilon$. As in the time domain, define

$$\begin{aligned} D_{NL}^{\dagger}\left(W_F\right) &= \frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^{*}\mathbf{M}_{W_F}W_{X,i}\right)-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\operatorname{Re}\left(W_{X,i}^{*}\mathbf{M}_{W_F}W_{X,k}a_{ik}\right) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{Z,i}\left(F\right)^{*}W_{Z,i}\left(F\right)\right), \end{aligned}$$

where $W_{Z,i}(F) = \mathbf{M}_{W_F}W_{X,i} - \frac{1}{N}\sum_{k=1}^{N}\mathbf{M}_{W_F}W_{X,k}a_{ik}$. Let $W_{Z,i} = W_{Z,i}(F^0)$ and $D_{NL}^{\dagger} = D_{NL}^{\dagger}(W_{F^0})$. Then we can define the variable Z_{it} like we did in time domain as if its DFT over Fourier frequencies is given by $W_{Z,i}$, and $\tilde{W}_{Z,i}$ is defined in the same manner as $\tilde{W}_{X,i}$ above. Let $f_{\varepsilon,i}(\cdot)$ denote the marginal spectral density of ε_{it}. We introduce some extra assumptions that are specified for the FDLS estimation together with some remarks.

Assumption A*. (i) Denote the $(P+R+1)\times 1$ vector $V_{it} = (X_{it}', F_t', \varepsilon_{it})'$. Suppose V_{it} is covariance stationary and has the spectral density matrix satisfying

$$f_{V,i}(\gamma) \sim \Gamma(\gamma)\Upsilon_i\Gamma(\gamma) \text{ as } \gamma \to 0^+, \tag{3.4.11}$$

where Υ_i is a $(P+R+1)\times(P+R+1)$ symmetric matrix that is finite uniformly over i with the following structure:

$$\Upsilon_i = \begin{pmatrix} \Upsilon_{i,XX} & \Upsilon_{i,XF} & 0 \\ \Upsilon'_{i,XF} & \Upsilon_{FF} & 0 \\ 0 & 0 & \Upsilon_{i,\varepsilon\varepsilon} \end{pmatrix},$$

in which for all i, the $P \times P$ and $R \times R$ submatrices $\Upsilon_{i,XX}$ and $\Upsilon_{i,FF}$ are positive definite, and the scalar $\Upsilon_{i,\varepsilon\varepsilon} > 0$. Γ is a diagonal matrix given by

$$\Gamma(\gamma) = Diag\left(\gamma^{-d_{X_1}}, \ldots, \gamma^{-d_{X_P}}, \gamma^{-d_{F_1}}, \ldots, \gamma^{-d_{F_R}}, \gamma^{-d_\varepsilon}\right).$$

(ii) There exists $\theta \in (0,2]$ such that for each i,

$$\left|f_{V,i,(ab)} - \upsilon_{i,(ab)}\gamma^{-d_a-d_b}\right| = O\left(\gamma^{\theta-d_a-d_b}\right) \text{ as } \gamma \to 0^+$$

for arbitrary $a, b = 1, \ldots, (P+R+1)$.

(iii) Let $V_{it} = \mu_V + \sum_{j=0}^{\infty} A_{V,j}\zeta_{V,i,t-j}$, where $A_{V,j}$ is a block-diagonal matrix consisting of $A_{X,j}$, $A_{F,j}$ and $A_{\varepsilon,j}$ in order, as given by (3.2.2)-(3.2.4). And define $A_V(\gamma) = \sum_{j=0}^{\infty} A_{V,j}e^{\mathbf{i}j\gamma}$. As $\gamma \to 0^+$,

$$\left\|\frac{dA_{V,a}(\gamma)}{d\gamma}\right\| = O\left(\gamma^{-1}\|A_{V,a}(\gamma)\|\right)$$

for arbitrary $a = 1, \ldots, (P+R+1)$, where $A_{V,a}(\gamma)$ is the a-th row of $A_{V,a}(\gamma)$.

Assumption B*. (i) Let $\check{\Gamma}_{X,j} = \text{Diag}\left(\gamma_j^{d_{X_p}}\right)$, we assume $E\left\|\check{\Gamma}_{X,j}W_{X,ij}\right\|^4 \leq M$ and $\frac{1}{T}\tilde{W}^*_{X,i}\tilde{W}_{X,i} \xrightarrow{p} \Sigma^W_{X,i} > 0$ for some matrix $\Sigma^W_{X,i}$, as $T \to \infty$ for each $i = 1, \ldots, N$.

(ii) Let $\mathcal{W} = \left\{\tilde{W}_F \in \mathbb{C}^{L\times R} : \tilde{W}_F = W_F\tilde{\Gamma}_F,\ \tilde{W}^*_F\tilde{W}_F/T = \mathbb{I}_R\right\}$. We assume

$\inf_{\tilde{W}\in\mathcal{W}} D^{\dagger}_{NL}\left(\tilde{W}\right) > 0$.

(iii) Let $\check{\Gamma}_{F,j} = \text{Diag}\left(\gamma_j^{d_{F_r}}\right)$, we assume $E\left\|\check{\Gamma}_{F,j}W_{F,j}\right\|^4 \leq M$ and $\frac{1}{T}\tilde{W}_F^*\tilde{W}_F \xrightarrow{p} \Sigma_F^W > 0$ for some matrix Σ_F^W, as $T \to \infty$.

Assumption C*. (i) $E\left\|\gamma_j^{d_\varepsilon}W_{\varepsilon,ij}\right\|^8 \leq M$ and $\frac{1}{T}\tilde{W}_{\varepsilon,i}^*\tilde{W}_{\varepsilon,i} \xrightarrow{p} \Sigma_{\varepsilon,i} > 0$ for some matrix $\Sigma_{\varepsilon,i}$, as $T \to \infty$.

(ii) Let $\sqrt{E\left|W_{\varepsilon,ik}W_{\varepsilon,jl}^*\right|^2} = \sigma_{ij,kl}^{W,1}$ and $E\left|W_{\varepsilon,ik}W_{\varepsilon,il}^*\right|^2 = \sigma_{i,kl}^{W,1}$. We assume $\sigma_{ij,kl}^{W,1} \leq \gamma_k^{-d_\varepsilon}\gamma_l^{-d_\varepsilon}\overline{\sigma}_{ij}^W$ and $\sqrt{\sigma_{i,kl}^{W,1}\sigma_{j,kl}^{W,1}} \leq \gamma_k^{-2d_\varepsilon}\gamma_l^{-2d_\varepsilon}\overline{\sigma}_{ij}^W$ for all (k,l) and

$$\frac{1}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij}^W \leq M. \tag{3.4.12}$$

Moreover, let $\left|E\left(W_{\varepsilon,ik}W_{\varepsilon,jl}^*\right)\right| = \sigma_{ij,kl}^{W,2}$. We assume $\left|\sigma_{ij,kl}^{W,2}\right| \leq \overline{\sigma}_{kl}^W$ for all (i,j), and

$$\frac{\gamma_L^{2d_\varepsilon}}{L^{1+2d_\varepsilon}}\sum_{k,l=1}^{L}\overline{\sigma}_{kl}^W \leq M,\ \frac{\gamma_L^{2d_\varepsilon}}{NL^{1+2d_\varepsilon}}\sum_{i,j=1}^{N}\sum_{k,l=1}^{L}\left|\sigma_{ij,kl}^W\right| \leq M,\text{ and } \frac{1}{T^{4d_\varepsilon}\log^2 L}\sum_{k,l=1}^{L}\gamma_N^W(k,l)^2 \leq M, \tag{3.4.13}$$

where $\gamma_N^W(k,l) = \frac{1}{N}\sum_{i=1}^N E\left(W_{\varepsilon,il}W_{\varepsilon,ik}^*\right)$.

(iii) Let $\Omega_i = \check{\Gamma}_\varepsilon E\left(W_{\varepsilon,i}W_{\varepsilon,i}^*\right)\check{\Gamma}_\varepsilon$, where $\check{\Gamma}_\varepsilon = \text{Diag}\left(\gamma_j^{d_\varepsilon}\right)$. The largest eigenvalue of Ω_i is bounded uniformly over i and T as $T \to \infty$.

(iv) For every (k,l), $E\left|N^{-\frac{1}{2}}\gamma_j^{d_\varepsilon}\gamma_l^{d_\varepsilon}\sum_{i=1}^N\left[W_{\varepsilon,ik}W_{\varepsilon,il}^* - E(W_{\varepsilon,ik}W_{\varepsilon,il}^*)\right]\right|^4 \leq M$.

(v) Moreover,

$$\frac{\gamma_L^{4d_\varepsilon}}{NL^2}\sum_{i,j=1}^{N}\sum_{k,l=1}^{L}\left|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,ik}^*, W_{\varepsilon,jl}W_{\varepsilon,jl}^*)\right| \leq M, \tag{3.4.14}$$

$$\frac{\gamma_L^{4d_\varepsilon}}{N^2L^2}\sum_{i,j,m,n=1}^{N}\sum_{k,l=1}^{L}\left|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,jk}^*, W_{\varepsilon,ml}W_{\varepsilon,nl}^*)\right| \leq M,$$

and

$$\frac{\gamma_L^{4d_\varepsilon}}{NL^2}\sum_{i,j=1}^{N}\sum_{k,l=1}^{L}\left|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,il}^*, W_{\varepsilon,jk}W_{\varepsilon,jl}^*)\right| \leq M.$$

Assumption D*. (i) Let $\Gamma_Z = Diag\left\{\gamma_L^{d_{Z_p}}\right\}$. $\text{plim}\,\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\left(W_{F^0}\right)\Gamma_Z = D_0^W$ for some matrix $D_0^W > 0$.

(ii) $\frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT}\sum_{i=1}^{N}\text{Re}\left(W_{Z,i}^* W_{\varepsilon,i}\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_0^W\right)$, where

$$\Sigma_0^W = \text{p}\lim \frac{\gamma_L^{2d_\varepsilon - 1}}{NT}\sum_{i,j=1}^{N}\text{Re}\left(\Gamma_Z W_{Z,i}^* W_{\varepsilon,i}\right)\text{Re}\left(W_{\varepsilon,i}^* W_{Z,i}\Gamma_Z\right) \equiv \text{p}\lim \Sigma_{NL}^{\dagger}.$$

(iii) Moreover, we have for each $i = 1, \ldots, N$ that

$$E\left\|\frac{\sqrt{L}\gamma_L^{d_\varepsilon - \frac{1}{2}}}{T}W_{\varepsilon,i}^*\tilde{W}_{X,i}\right\|^2 \le M,\ E\left\|\frac{\sqrt{L}\gamma_L^{d_\varepsilon - \frac{1}{2}}}{T}W_{\varepsilon,i}^*\tilde{W}_F\right\|^2 \le M,$$

and

$$E\left\|\frac{\sqrt{L}\gamma_L^{d_\varepsilon - \frac{1}{2}}}{T}W_{\varepsilon,i}^*\tilde{W}_{Z,i}\right\|^2 \le M,$$

and the following also holds:

$$E\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon - \frac{1}{2}}}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^*\tilde{W}_{X,i}\right\|^2 \le M,$$

$$E\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon - \frac{1}{2}}}{NT}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^*\tilde{W}_F\right\|^2 \le M, \text{ and } E\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon - \frac{1}{2}}}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^*\tilde{W}_F\right\|^2 \le M.$$

Assumption E*. As $T \to \infty$, we assume (i) $\frac{L}{T} + \frac{1}{L} \to 0$. Moreover, denote $\overline{d} = \max\left\{d_{X,\max}, d_{F,\max}, d_{Z,\max}\right\}$, $\underline{d} = \min\left\{d_{X,\min}, d_{F,\min}, d_{Z,\min}\right\}$, and $\Delta d = \overline{d} - \underline{d}$, then we assume:

(ii) $d_\varepsilon < \frac{1}{4}$, $\underline{d} > d_\varepsilon$ and $7\left(\frac{1}{2} - \overline{d}\right) > \frac{1}{2}$; and (iii) $\left(\frac{1}{N^{\frac{1}{3}}} + \sqrt{\frac{L}{N}} + \frac{1}{\sqrt{L}}\right)\gamma_L^{-2\Delta d} \to 0$.

Remark 7. Assumption A* is basically the standard restrictions of multivariate stationary long memory processes (see, e.g. Christensen and Nielsen, 2006). Assumption A*(i) complements Assumption A(i) as it defines the long memory processes through their joint spectral density matrix around zero frequencies, where a certain power law is satisfied. Under stationary long memory, Assumption A*(i) and A(i) are basically equivalent to each other, but the former assumption in frequency domain also holds uniformly under short memory. Also note that we allow the spectral density to have heterogeneous constant multiplier across i, which indicates a heterogeneous cross-correlation and auto-correlation structure among different cross-sectional units.

Remark 8. Assumption B* extends the conditions in Assumption B to the frequency domain, as it is consistent with the result of probabilistic limit for the average (cross-) periodogram, which is shown in Theorem 1 of Robinson (1994b) for univariate case and Theorem 1 in Lobato (1997) for multivariate case. It is also consistent with the results on how the expectation of periodogram approximates the spectral density as indicated by (3.16) in the proof of Theorem 1 in Robinson (1995a).

Remark 9. Assumption C* extends Assumption C to the frequency domain, where C*(i) gives the probabilistic limit of averaged periodogram and the fourth order moment of periodogram of the idiosyncratic error as in B*(i) and B*(iii). And C*(ii) gives the conditions of cross-sectional weak dependence in (3.4.12) and of serial dependence over frequencies in (3.4.13) for the DFT of idiosyncratic error. Conditions given by (3.4.12) is slightly stronger than those analogs in (3.3.2) in time domain, and we impose them so as to support C*(v). The conditions given by (3.4.13) adopt the Theorem 2 in Robinson (1995b) which gives the limit of expectation of $W^*_{\varepsilon,i}W_{\varepsilon,i}$ at Fourier frequencies defined above. To see how it holds, we use the fact that $\left|E\left(W_{\varepsilon,ik}W^*_{\varepsilon,jl}\right)\right| \leq \sqrt{E\left(W_{\varepsilon,ik}W^*_{\varepsilon,jl}\right)^2}$ and call upon Theorem 1 in Robinson (1995b), which in our setting indicates that

$$\lim_{T\to\infty} E\left\{\frac{W_{\varepsilon,ik}W^*_{\varepsilon,il}}{f_{\varepsilon,i}\left(\gamma_k\right)^{\frac{1}{2}} f_{\varepsilon,i}\left(\gamma_l\right)^{\frac{1}{2}}}\right\} = P_d\left(k,l\right), \text{ with } P_d\left(k,l\right) \leq M\frac{(kl)^{d_\varepsilon}}{k+l}$$

for some positive constant $M < \infty$. Then

$$\begin{aligned}
\sum_{k,l=1}^{L} \sigma_{i,kl}^{W} &\leq M \sum_{k,l=1}^{L} \frac{(kl)^{d_\varepsilon}}{k+l} f_{\varepsilon,i}(\gamma_k)^{\frac{1}{2}} f_{\varepsilon,i}(\gamma_l)^{\frac{1}{2}} \\
&\sim M \sum_{k,l=1}^{L} \frac{(kl)^{d_\varepsilon}}{k+l} \gamma_k^{-d_\varepsilon} \gamma_l^{-d_\varepsilon} \leq M \left(\sum_{k=1}^{L} k^{d_\varepsilon - \frac{1}{2}} \gamma_k^{-d_\varepsilon} \right)^2 \\
&= \frac{M}{(2\pi)^{2d_\varepsilon - 1}} \left(T^{d_\varepsilon + \frac{1}{2}} \frac{1}{T} \sum_{k=1}^{L} \gamma_k^{-\frac{1}{2}} \right)^2 \\
&\approx \frac{M}{(2\pi)^{2d_\varepsilon - 1}} \left(T^{d_\varepsilon + \frac{1}{2}} \gamma_L^{\frac{1}{2}} \left\{ 1 + O\left(\frac{1}{T}\right) \right\} \right)^2 = O\left(\gamma_L^{-2d_\varepsilon} L^{1+2d_\varepsilon} \right),
\end{aligned}$$

which explains the orders in (3.4.12). By the same reasoning, (3.4.13) can be explained through

$$\begin{aligned}
\sum_{k,l=1}^{L} \gamma_N^W (k,l)^2 &\leq M \sum_{k,l=1}^{L} \frac{(kl)^{2d_\varepsilon}}{(k+l)^2} f_{\varepsilon,i}(\gamma_k) f_{\varepsilon,i}(\gamma_l) \\
&\leq M \left(\sum_{k=1}^{L} k^{2d_\varepsilon - 1} \gamma_k^{-2d_\varepsilon} \right)^2 \leq M \left(T^{2d_\varepsilon} \sum_{k=1}^{L} k^{-1} \right)^2 = O\left(T^{4d_\varepsilon} \log^2 L \right).
\end{aligned}$$

C*(iii) mimics the Assumption C(ii) in Bai (2009) in time domain, as it adopts C*(i) to control the order in frequency. And C*(iv) continues to illustrate the weak cross-sectional dependence.

Remark 10. And C*(v) gives some higher order conditions that mimic the ones in time domain setup. To give more explanation, denote $\check{W}_{\varepsilon,ij} = \gamma_j^{d_\varepsilon} W_{\varepsilon,ij}$. And take (3.4.14) for instance, we have

$$\begin{aligned}
&\sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \left| \mathrm{cov}(W_{\varepsilon,ik} W_{\varepsilon,ik}^{*}, W_{\varepsilon,jl} W_{\varepsilon,jl}^{*}) \right| \\
&= \sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \left| \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \mathrm{cov}(\check{W}_{\varepsilon,ik} \check{W}_{\varepsilon,ik}^{*}, \check{W}_{\varepsilon,jl} \check{W}_{\varepsilon,jl}^{*}) \right| \\
&\leq \sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \sqrt{\mathrm{Var}\left(\check{W}_{\varepsilon,ik} \check{W}_{\varepsilon,ik}^{*} \right) \mathrm{Var}\left(\check{W}_{\varepsilon,jl} \check{W}_{\varepsilon,jl}^{*} \right)} \\
&\leq \sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \sqrt{E \left| \check{W}_{\varepsilon,ik} \check{W}_{\varepsilon,ik}^{*} \right|^2 E \left| \check{W}_{\varepsilon,jl} \check{W}_{\varepsilon,jl}^{*} \right|^2}
\end{aligned}$$

$$\leq \sum_{k,l=1}^{L} \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \sum_{i,j=1}^{N} \sqrt{\overline{\sigma}_i^{W,2} \overline{\sigma}_j^{W,2}} = O\left(NT^2 \gamma_L^{2-4d_\varepsilon}\right)$$
$$= O\left(NL^2 \gamma_L^{-4d_\varepsilon}\right)$$

using the second inequality in (3.4.12) and Riemann sum approximation.

Remark 11. Assumption D*(i) and D*(ii) extends the distributional theory in Theorem 2 of Christensen and Nielsen (2006) into factor model structure. This is relatively high-order assumption and is not a trivial result because $W_{Z,i}$ by construction is not the DFT of a linear process like $W_{X,i}$ and W_F. Note that different from time domain setup, Assumption C* does not really impose the "weak dependence" over frequencies as the normalization there is only slightly stronger than using the limit of averaged periodogram. Such conclusion is also confirmed by the proof of Theorem 2 in Christensen and Nielsen (2006) as weak dependence over frequencies only occurs in cross-periodogram between the error and the regressors, rather than the periodograms, and this conclusion is reflected by Assumption D*(iii) in our factor model.

In the following we present some asymptotic theoretical results of our FDLS estimator. Then in the following proposition we firstly establish the result of consistency for FDLS estimators $\tilde{\beta}$ and $\hat{W}_F$.

Proposition 3.4.1 *Suppose Assumptions A-D and A*-B* hold. Then as* $(N, T) \to \infty$ *we have*

(i) The FDLS estimator β *is consistent as* $\tilde{\beta} \xrightarrow{p} \beta$*;*

(ii) The matrix $\tilde{W}_{F^0}^* \hat{W}_F / T$ *is invertible and* $\left\| \mathbf{P}_{\hat{W}_F} - \mathbf{P}_{W_{F^0}} \right\| \xrightarrow{p} 0$.

In the above proposition, (i) establishes the consistency of $\hat{\beta}^W$ and(ii) indicates that the columns of $\hat{W}_F$ span the same space as those of $\tilde{W}_{F^0}$ asymptotically. These results are intermediately used in the subsequent analysis.

Next, we establish the theory of asymptotic distribution of $\hat{\beta}^W$ and its asymptotic bias terms, given by the following theorem.

Theorem 3.4.2 *Suppose Assumptions A, B and A*-E* hold. Then as* $(N,T) \to \infty$ *for comparable* N *and* T *such that* $N/T \to \rho$*, we have*

$$\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0-A_{NT}^W\right)\xrightarrow{d}\mathcal{N}\left(0,\left(D_0^W\right)^{-1}\Sigma_0^W\left(D_0^W\right)^{-1}\right),$$

for positive definite matrices D_0^W *and* Σ_0^W *defined in Assumption D*, where* $A_{NT}^W = O_p(\phi_L)$ *with*

$$\phi_L = \frac{\gamma_L^{2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}}{L}.$$

The asymptotic bias term A_{NT}^W *is given by*

$$\begin{aligned}A_{NT}^W &= -\Gamma_Z\left(D_{NL}^W\right)^{-1}\Gamma_Z\frac{\gamma_L^{-1}}{NT}\sum_{i=1}^N Re\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^N Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\tilde{G}\tilde{\lambda}_i\right)\\ &= -\left(D_{NL}^\dagger\left(W_{F^0}\right)\right)^{-1}\frac{1}{NT}\sum_{i=1}^N Re\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^N Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\tilde{G}\tilde{\lambda}_i\right)\end{aligned}$$

where $\tilde{G} = \left(\frac{\tilde{W}_{F0}^*\hat{W}_F}{T}\right)^{-1}\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}$*, and* $D_{NL}^W = \gamma_L^{-1}\Gamma_Z D_{NL}^\dagger(W_{F^0})\Gamma_Z = O_p(1)$.

Note that in frequency domain, L (or $h=\frac{L}{T}$) can be treated as a "bandwidth" as it measures the width of frequency region local to zero. Then instead of dealing with summation over both dimensions as in a classic panel model, our asymptotic theory is more like the one for a cross-sectional nonparametric regression model, where the limiting theory can be obtained by large N alone, and the large T helps to control the bandwidth. This insight can be partly explain by the component $\sqrt{NL}=\sqrt{NTh}$ in our convergence rate. By this reasoning we need $L=O(T^\alpha)$, with $0<\alpha<1$, which can obtains an elegant form of asymptotic distribution as shown above, but at the cost of efficiency. This reflects the bias-variance trade-off in the nonparametric model studies.

And in terms of the bias, we have only one nonnegligible term that correspond to the one with order $\frac{1}{T}$ in time domain. This is because the sample size in frequency domain is of smaller order of T, and thus of N when N

and T have comparable size as we assume. But here we still keep it explicit in our asymptotic distribution so as to help compare with time domain least squares estimator in this paper and in Bai (2009). We can correct these bias using half-panel jackknife as we did in the time domain setting, which require estimation of all memory parameters as before. But after taking DFT, over frequency domain we have asymptotic serial uncorrelation in large T as indicated by Assumption C*. This implies we have weak dependence on both cross-sectional units and frequencies now, therefore we can consider estimate the bias term with its analytical form in frequency domain analogous to the ones in Bai (2009, pp. 1250). Specifically, we consider directly estimating

$$\bar{A}^W_{NT} = -\frac{1}{NT} D^{\dagger}_{NL} \left(W_{F^0}\right)^{-1} \sum_{i=1}^{N} \mathrm{Re}\left(W^*_{X,i} \mathbf{M}_{\hat{W}_F} \frac{1}{N} \sum_{k=1}^{N} Diag\left(|W_{\varepsilon,kj}|^2\right) \hat{W}_F \tilde{G} \tilde{\lambda}_i \right)$$

by replacing W_{F^0} with $\hat{W}_F$ and $\tilde{\lambda}_i$ with $\hat{\lambda}_i$ from (3.4.7). In addition, we can replace $Diag\left(|W_{\varepsilon,kj}|^2\right)$ by $\hat{\Omega}_k = Diag\left(\hat{W}_{\varepsilon,k1}\hat{W}^*_{\varepsilon,k1}, \ldots, \hat{W}_{\varepsilon,kL}\hat{W}^*_{\varepsilon,kL}\right)$, where $\hat{W}_{\varepsilon,i} = W_{Y,i} - \hat{W}_{X,i}\tilde{\beta} - \hat{W}_F\hat{\lambda}_i$. That is, we estimate $\bar{A}_{NT}$ by

$$\hat{A}^W_{NT} = -\frac{1}{NT}\left(\hat{D}^W_{NL}\right)^{-1} \sum_{i=1}^{N} \mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F} \frac{1}{NT} \sum_{k=1}^{N} \hat{\Omega}_k \hat{W}_F \left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1} \hat{\lambda}_i \right), \tag{3.4.15}$$

where

$$\hat{D}^W_{NL} = \frac{1}{NT}\sum_{i=1}^{N} \mathrm{Re}\left(\hat{W}^*_{Z,i}\hat{W}_{Z,i}\right),$$

with $\hat{W}_{Z,i} = W^*_{X,i}\mathbf{M}_{\hat{W}_F} - \frac{1}{N}\sum_{k=1}^{N} W^*_{X,k}\mathbf{M}_{\hat{W}_F}\hat{a}_{ik}$ and $\hat{a}_{ik} = \hat{\lambda}'_i \left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_k$ estimated by our PCA in frequency domain. Then the bias-corrected estimator is given by $\tilde{\beta}^{bc} = \tilde{\beta} - \hat{A}^W_{NT}$.

To proceed, we consider a feasible implementation of the inference given by Theorem 3.4.2, where we do not need to estimate any of the memory parameters. To do that, we estimate the asymptotic covariance matrix using $\left(\hat{D}^W_{NL}\right)^{-1} \hat{\Sigma}^W_{NL} \left(\hat{D}^W_{NL}\right)^{-1}$, where for $\hat{\Sigma}^W_{NL}$ we focus on the case when $\zeta_{\varepsilon,i,t}$ is independent across i, and the estimator is given by

$$\hat{\Sigma}_{NL}^{W}=\frac{1}{N^2T^2}\sum_{i=1}^{N}\mathrm{Re}\left(\hat{W}_{Z,i}^{*}\hat{W}_{\varepsilon,i}\right)\mathrm{Re}\left(\hat{W}_{\varepsilon,i}^{*}\hat{W}_{Z,i}\right)$$

The following theorem establishes the asymptotic distribution of $\tilde{\beta}^{bc}$ using the above estimation of covariance matrix.

Theorem 3.4.3 *Suppose Assumptions A, B and Assumption A^*-E^* hold; and as $N,T\to\infty$ for comparable N and T such that $N/T\to\rho$ we have*

$$\left(\hat{\Sigma}_{NL}^{W}\right)^{-\frac{1}{2}}\hat{D}_{NL}^{W}\left(\tilde{\beta}^{bc}-\beta^{0}\right)\xrightarrow{d}N\left(0,\mathbb{I}_P\right).$$

The above theorem adopts the idea of self-normalization. To see this, since $\tilde{\beta}^{bc}$ consistently correct the bias, we have from above that

$$\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}^{bc}-\beta^{0}\right)\xrightarrow{d}\mathcal{N}\left(0,\left(D_0^W\right)^{-1}\Sigma_0^W\left(D_0^W\right)^{-1}\right),$$

which includes memory parameters that were treated as coefficients to be estimated. The inference of $\tilde{\beta}$ based on the plug-in estimator (like local Whittle) of these memory parameters could perform not that well in finite sample. So we try to fix it by showing

$$\left(\hat{D}_{NL}^{W}\right)^{-1}\hat{\Sigma}_{NL}^{W}\left(\hat{D}_{NL}^{W}\right)^{-1}\xrightarrow{p}\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(D_0^W\right)^{-1}\Sigma_0^W\left(D_0^W\right)^{-1}\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}$$

which is presented in the proof of Theorem 3.4.3.

3.5 Determination of Number of Factors

So far we have assumed to know the true number of factors in analysis of model (3.2.1), but in practice one has to determine the number of factors to use. This leads to the requirement of consistently estimating the true number of factors (denoted by R^0 afterwards), which is an intrinsic issue in factor analysis. For pure factor model, there are various methods to determine R^0; see Bai and Ng (2002) for an information criterion approach, and Onatski (2010) for an "edge distribution" approach, and Ahn and Horenstein (2013) (2013, AH afterwards) for eigenvalue ratio and growth ratio approach. In this section we try to modify the eigenvalue ratio (ER) estimation into our regression model in frequency domain, and illustrate its consistency. To proceed, we specify an upper bound $R_{\max} \geq R^0$. The procedure goes as the following:

1. Conduct the frequency domain least squares using $R_{\max}$ factors, as indicated by (3.4.9) and (3.4.10) using model (3.4.4), and correct its bias by Theorem 3.4.3. Denote this estimator by $\tilde{\beta}_{(R_{\max})}$.

2. Let $\widetilde{u}_{it} = Y_{it} - X_{it}'\tilde{\beta}_{(R_{\max})}$ be the "partial" residual of FDLS in time domain, which builds up a $N \times T$ real matrix $\widetilde{U}$. Then derive the first $R_{\max} + 1$-th largest eigenvalues of $\widetilde{U}\widetilde{U}'/NT$, denoted by $\tilde{\mu}_{NT,j}$, $j = 1, \ldots, R_{\max} + 1$.

3. Let the eigenvalue ratio be $ER\left(j\right) = \frac{\tilde{\mu}_{NT,j}}{\tilde{\mu}_{NT,j+1}}$ and ER estimate of number of factors is given by $\widetilde{r}_{ER} = \max_{1 \leq j \leq R_{\max}} ER\left(j\right)$.

The above method is a modification of ER estimate from AH as we have to take the regression structure into account. By the reasoning as in Bai (2009), $\widetilde{u}_{it}$ follows a pure factor model approximately, and thus ER estimate can be applied. To be exact,

$$\begin{aligned}
\widetilde{u}_{it} &= Y_{it} - X_{it}'\tilde{\beta}_{(R_{\max})} = X_{it}'\left(\beta - \tilde{\beta}_{(R_{\max})}\right) + u_{it} \\
&= X_{it}'\left(\beta - \tilde{\beta}_{(R_{\max})}\right) + \lambda_i' F_t + \varepsilon_{it}. \qquad (3.5.1)
\end{aligned}$$

And considering the complex convergence rate and bias order pattern of time domain least squares estimator, as presented in Theorem 3.3.1, we use frequency domain estimator in step 1 above, which makes this procedure a hybrid approach of time and frequency domain. To proceed, we define some additional notations adopted from AH. Let $\psi_k(A)$ be the k-th largest eigenvalue of matrix A; and $m = \min(N,T)$ and $M = \max(N,T)$; and $\lfloor \cdot \rfloor$ the integer part of a number. In the following we present the consistency of our modified ER estimate with some extra assumptions imposed.

Assumption G. (i) Let $\mu_{NT,k} = \psi_k\left[\left(\Lambda'\Lambda/N\right)\left(F'F/T\right)\right]$ for $k = 1, \ldots, R^0$. Then $\operatorname{plim} \mu_{NT,k} = \mu_k$ for some $\mu_k \in (0, \infty)$ and for each $k = 1, \ldots R^0$. (ii) R^0 is finite.

Assumption H. (i) $0 < y \equiv \lim_{m\to\infty} \frac{m}{M} \leq 1$.

(ii) Let E be the $N \times T$ matrix consisting of ε_{it}, then $E = R_T^{\frac{1}{2}} Z G_N^{\frac{1}{2}}$ where Z is an $N \times T$ matrix with i.i.d. element along both dimensions with finite fourth moment; and $R_T^{\frac{1}{2}}$ and $G_N^{\frac{1}{2}}$ are symmetric square roots of positive definite matrices $R_T : T \times T$ and $G_N : N \times N$ with $\psi_1(R_T) < c_1$, $\psi_1(G_N) < c_1$ uniformly over N and T respectively.

Assumption I. (i) $\psi_T(R_T) > c_2$ for all T.

(ii) Let $y^* = \lim_{m\to\infty} \frac{m}{N} = \min(y, 1)$. Then there exists a real number $d^* \in (1 - y^*, 1]$ such that $\psi_{\lfloor d^* N \rfloor}(G_N) > c_2$ for all N.

Assumption J. Consider linear combinations $W_X \cdot \alpha \equiv \sum_{p=1}^{P} \alpha_p W_X^p$ such that W_X^p is an $N \times L$ complex matrix of DFT of the p-th element of regressor, and the $P \times 1$ vector α satisfies $\|\alpha\| = 1$. There exists a constant $b > 0$ such that

$$\min_{\alpha \in \mathbb{R}^P, \|\alpha\|=1} \frac{1}{NT} \sum_{r=R+R^0+1}^{L} \mu_r\left[(W_X \cdot \alpha)^* (W_X \cdot \alpha)\right] \geq b \text{ w.p.a.1.}$$

The Assumption G-I are directly borrowed from the Assumption A, C and D in AH. These three assumptions are not related to the level of persistence among any variables, and thus can continue to hold under our setup. Assumption B in AH gives the moment conditions and cross-sectional and serial dependence of factors, factor loadings and idiosyncratic error, which is already covered by our Assumption B and C and thus is compatible with long memory. As we show in the proof of Theorem 3.5.1 in Appendix, the conditions of serial dependence do not play an important role in proving the consistency of $\widetilde{r}_{ER}$, so most of the proof of Theorem 1 in AH and relevant lemmas continue to hold. Assumption J is the frequency domain extension of Assumption NC in Moon and Weidner (2015), which also illustrates the noncollinearity of regressor. With the above extra assumptions, we establish the following theorem:

Theorem 3.5.1 *Suppose Assumption B, C, G-J and B^*, C^* and E^* hold with $R^0 \geq 1$, we have $\left\| \tilde{\beta}_{(R_{\max})} - \beta \right\| = O_p\left(\gamma_L^{\frac{1}{2} - d_{X,\max}} \right)$; and there exists $d^c \in (0, 1]$ such that $\lim_{m\to\infty} \Pr(\widetilde{r}_{ER} = R^0) = 1$ for any $R_{\max} \in (R^0, \lfloor d^c m \rfloor - R^0 - 1]$.*

3.6 Monte Carlo Simulations

3.6.1 Results using time domain least squares estimation

Firstly we examine the performance of time domain LS estimator in finite sample, using the following model

$$Y_{it} = X_{it}'\beta^0 + \lambda_i' F_t^0 + \varepsilon_{it} \tag{3.6.1}$$

where $\beta^0 = (0.6, 0.9)'$. Moreover, we consider four sets of data generating processes (DGP hereafter) with two regressors ($K = 2$) and two factors ($R^0 = 2$). In detail, the idiosyncratic error is firstly generated by

$$\varepsilon_{it} = 0.4\varepsilon_{i,t-1} + e_{it}; \tag{3.6.2}$$

and we consider the cases with and without conditional heteroskedasticity for ε_{it}. When idiosyncratic error term is conditionally heteroskedastic, we adopt the multiplier as

$$\varepsilon_{it}^{*}=0.06\sqrt{\frac{X_{it}^{\prime}X_{it}}{K}}\varepsilon_{it}. \tag{3.6.3}$$

The above setup characterizes two varieties of DGP. And the other two are specified by the following DGPs of factor:

$$F_t^0=2-0.7F_{t-1}^0+e_{f,t}, \tag{3.6.4}$$

and

$$F_t^0=2-0.2F_{t-1}^0+e_{f,t}, \tag{3.6.5}$$

where $e_{f,t}\sim I\left(d_f\right)$ and $e_{it}\sim I\left(d_e\right)$ are two (multivariate) fractional integrated processes generated by i.i.d. $t\left(5\right)$ innovations. In terms of regressor, it is generated by

$$X_{it}=\sum_{r=1}^{2}\left(\chi_{ri}+\lambda_{ri}\right)\left(F_{rt}^0+F_{r,t-1}^0\right)+\widetilde{X}_{it} \tag{3.6.6}$$

with $\widetilde{X}_{it}=1-0.8\widetilde{X}_{it-1}+x_{it}$. Four DGPs we try to specify are denoted by DGP1 to DGP4, where DGP1 combines (3.6.2) and (3.6.4), DGP2 combines (3.6.3) and (3.6.4), DGP3 combines (3.6.2) and (3.6.5), and DGP4 combines (3.6.3) and (3.6.5). Random variables f_{rt}, e_{it}, x_{it}, λ_{ri}, and χ_{ri} are mutually independent, and $x_{it}\sim I\left(d_X\right)$ with i.i.d. $t\left(5\right)$ innovations as above, and χ_{ri} and $\lambda_{ri}\sim\mathcal{N}\left(1,1\right)$. As a benchmark, we firstly estimate β using LS estimator with its bias corrected by the classic half-panel jackknife using the bias order $\frac{1}{N}+\frac{1}{T}$ as in Fernández-Val and Weidner (2016), denoted by $\hat{\beta}$. To evaluate the performance of LS estimator, we choose the first argument of β and present its Monte Carlo root mean squared error (RMSE) of $\hat{\beta}-\beta$ and $\left\|P_{\hat{F}}-P_{F^0}\right\|_F$ with projection matrix $P_A=A\left(A'A\right)^{-1}A'$. Also we present Monte Carlo standard deviation (STD) of $\hat{\beta}-\beta$ to examine the order of its variance, and by comparing RMSE and STD, we can see if the bias is still prominent or corrected properly by classic half-panel jackknife method. Moreover, to illustrate the asymptotic normality, we present 95%-level empirical coverage probability

(COVP) $\Pr\left(\left|\frac{\hat{\beta}_k-\beta_k}{\widehat{\sigma}_k}\right| \leq Z_{0.025}\right)$ with standard normal critical values denoted by Z, and for some cases we compare the histogram of $\frac{\hat{\beta}-\beta}{\widehat{\sigma}}$ against standard normality for illustration of its asymptotic distribution, where $\widehat{\sigma}$ is the Monte Carlo standard deviation of $\hat{\beta}$. We repeat the simulation with 200 repetitions, with sample size $N = T = 100$ and 200.

To cover as many cases within the range of short memory and stationary long memory, we present the results of following combinations of memory parameters. Firstly we consider the short memory case $d_f = d_e = d_X = 0$. And then for long memory cases, we consider $d_f = d_X = 0.2$ and $d_e = 0.1$; and $d_f = d_X = 0.2$ and $d_e = 0.3$; and $d_f = d_X = d_e = 0.3$; and $d_f = d_X = 0.4$ and $d_e = 0.3$. Note that our selection include the cases of short memory, weakly and strong long memory, and when fractional cointegration condition holds and does not hold. Results are presented in left panel of Table 1. And in terms of bias correction in time domain, Theorem 3.3.1 shows that only d_e enters the order of bias when factors and regressors have nonzero mean. We examine this theory by considering the bias-corrected estimator $\tilde{\beta}$ defined by

$$\tilde{\beta} = \hat{\beta} - \frac{1}{T^{1-2d_e}}C - \frac{1}{N}B,$$

given the true memory parameter d_e, which is an infeasible estimator. And we also present the performance of feasible bias-corrected estimator $\tilde{\beta}^*$, which is an adjusted half-panel jackknife bias correction by memory parameter, which is given by

$$\tilde{\beta}^* = \left(2 + \frac{1}{2^{1-2\widehat{d}_{\widehat{\varepsilon}}} - 1}\right)\hat{\beta} - \tilde{\beta}^*_{N/2,T} - \frac{1}{2^{1-2\widehat{d}_{\widehat{\varepsilon}}} - 1}\tilde{\beta}^*_{N,T/2},$$

where $\widehat{d}_{\widehat{\varepsilon}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{d}_{\widehat{\varepsilon}_i}$ is the individual average of the local Whittle estimator of regression residuals. The estimate $\tilde{\beta}^*_{N/2,T}$ is defined by average of LS estimator using half-panels given respectively by $\{i = 1, \ldots, \lceil N/2 \rceil; t = 1, \ldots, T\}$ and $\{i = \lfloor N/2 \rfloor + 1, \ldots, N; t = 1, \ldots, T\}$, with $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ being the ceiling and floor functions; and $\tilde{\beta}^*_{k,N,T/2}$ is defined similarly using half-panels given by $\{i = 1, \ldots, N; t = 1, \ldots, \lceil T/2 \rceil\}$ and $\{i = 1, \ldots, N; t = \lfloor T/2 \rfloor + 1, \ldots, T\}$. To intuitively see how $\tilde{\beta}^*$ correct the bias terms, we rewrite it as

$$\tilde{\beta}^* = \hat{\beta} - \left(\tilde{\beta}^*_{N/2,T} - \hat{\beta} \right) - \frac{1}{2^{1-2\hat{d}_{\hat{\varepsilon}}} - 1} \left(\tilde{\beta}^*_{N,T/2} - \hat{\beta} \right).$$

Note that by Theorem 3.3.1,

$$\hat{\beta} = \beta^0 + \frac{1}{T^{1-2d_\varepsilon}} C + \frac{1}{N} B,$$

then in asymptotics we have

$$\tilde{\beta}^*_{k,N/2,T} = \beta^0 + \frac{1}{T^{1-2d_\varepsilon}} C + \frac{2}{N} B, \text{ and } \tilde{\beta}^*_{k,N,T/2} = \beta^0 + \frac{2^{1-2d_\varepsilon}}{T^{1-2d_\varepsilon}} C + \frac{1}{N} B.$$

Therefore $\tilde{\beta}^*_{k,N/2,T} - \hat{\beta}_k = \frac{1}{N} B$ and $\tilde{\beta}^*_{k,N,T/2} - \hat{\beta}_k = \frac{\left(2^{1-2d_\varepsilon} - 1\right)}{T^{1-2d_\varepsilon}} C$, and by substitution we can see how $\tilde{\beta}^*_k$ correct the bias. Results of the above bias correction are presented in Table 3, where we consider a benchmark setup as we set both factor and regressor short memory variables.

From left panel of Table 1 and 2, we can see that time domain least squares estimator has prominent downward bias issue, especially when the joint memory is strong enough, which affects the inference as our coverage probabilities show. In Table 3, although in different setup, we can still see how bias is well corrected comparing the bias and coverage probability of $\hat{\beta}$ with that of $\tilde{\beta}^*$. But the results of feasible estimator $\tilde{\beta}^*_F$, where we replace all the memory parameters with its local Whittle-based estimator, is not good in some cases, which shows the necessity of using a more valid method as our frequency domain least squares estimation in the following.

3.6.2 Results using frequency domain least squares estimation

To examine the finite sample performance of frequency domain LS estimator, we consider the same data generating process as in time domain case above, given by (3.6.1)-(3.6.6). And we conduct DFT to (3.6.1) over the frequencies $\gamma_j = \frac{2\pi j}{T}$ for $j = 1, \ldots, L$ where $L = \left\lfloor T^{\frac{4}{5}} \right\rfloor$. This implies the model

$$W_{Y,ij} = W'_{X,ij}\beta + \lambda_i' W_{F,j} + W_{\varepsilon,ij},\ i = 1, \ldots, N,\ j = 1, \ldots, \left\lfloor T^{\frac{4}{5}} \right\rfloor$$

as (3.4.2). We present performance of $\tilde{\beta}_k^{**}$, the bias-corrected estimator given by Theorem 3.4.3, $k = 1, 2$, indicated by (3.4.9) and (3.4.10), including the root mean square error (RMSE), the standard deviation (STD), the 95%-level empirical coverage probablity (COVP) Pr $\left(\left| \frac{\tilde{\beta}_k^{**} - \beta_k}{\widehat{\sigma}_k} \right| \leq Z_{0.025} \right)$ with $\widehat{\sigma}_k$ the estimation of asymptotic variance of $\tilde{\beta}_k^{**}$ given by 3.4.3 as well, and the bias (BIAS), all averaged over repetitions. And to thoroughly compare the FDLS with least squares estimator in time domain, the above results are presented in right panel of Table 1 and 2. We can see that under short memory and relatively weak long memory case, time domain least squares estimator does not outperform the results of FDLS in a great deal, as they both obtain good bias control and coverage probabilities, although FDLS is a bit less efficient due to loss in sample size after DFT. However, when joint memory is relatively stronger, FDLS performs better as it prominently correct the bias and obtains a good coverage probability in most cases.

In the end we adopt the estimation of true number of factors introduced in last section, where we consider DGP1, but with true number of factors R^0 equal to either 2 or 3. We specify $R_{\max} = 8$ for both cases, and for the estimated number of factors over repetitions, we report the average (Mean), median (Median), ratio of correct estimation (RCE), over-estimation (ROE) and under-estimation (RDE) of the true number R^0. In Table 4 we presents the above results together with using the method of information criterion (IC) function proposed by Bai (2009), which is given by

$$IC_{p1}\left(k\right) = \ln\left(V\left(k, \hat{F}^k\right)\right) + k\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

where $V\left(k, \hat{F}^k\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}^2$, with $\widehat{\varepsilon}_{it}^2$ the residual given by time domain least squares estimator. From Table 4, we can see that in asymptotics, ER estimate outperform the IC one, especially when joint memory is strong enough. The only concern is that under relatively small sample, ER may suffer from the issue of under-estimating the number of factor, which may lead to an inconsistent least squares estimator based on that estimated R.

Table 1: LS estimate over frequency domain and its bias correction (DGP1 and 2)

DGP1								
$\hat{\beta}$ (bias-corrected time domain LS estimate by Bai (2009))					$\tilde{\beta}^{**}_{(A)}$ (bias-corrected FDLS estimate)			
	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)
$d_e = 0,\ d_f = 0,\ d_X = 0$								
$N, T = 100$	-0.000	0.004	0.004	0.960	-0.001	0.012	0.012	0.938
$N, T = 200$	0.000	0.002	0.002	0.965	-0.001	0.008	0.008	0.960
$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$								
$N, T = 100$	-0.002	0.007	0.007	0.940	-0.003	0.014	0.014	0.925
$N, T = 200$	-0.001	0.003	0.003	0.950	-0.002	0.010	0.010	0.940
$d_e = 0.3,\ d_f = 0.3,\ d_X = 0.3$								
$N, T = 100$	-0.012	0.008	0.014	0.680	-0.005	0.021	0.021	0.948
$N, T = 200$	-0.009	0.005	0.010	0.565	-0.004	0.014	0.014	0.953
$d_e = 0.3,\ d_f = 0.4,\ d_X = 0.4$								
$N, T = 100$	-0.011	0.009	0.015	0.800	-0.010	0.022	0.024	0.928
$N, T = 200$	-0.009	0.005	0.011	0.600	-0.005	0.013	0.014	0.920
DGP2								
$\hat{\beta}$ (bias-corrected time domain LS estimate by Bai (2009))					$\tilde{\beta}^{**}_{(A)}$ (bias-corrected FDLS estimate)			
	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)
$d_e = 0,\ d_f = 0,\ d_X = 0$								
$N, T = 100$	-0.001	0.009	0.009	0.965	0.000	0.013	0.013	0.945
$N, T = 200$	-0.000	0.004	0.004	0.950	0.000	0.006	0.006	0.963
$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$								
$N, T = 100$	-0.001	0.009	0.009	0.955	-0.004	0.014	0.015	0.940
$N, T = 200$	-0.001	0.005	0.005	0.955	-0.001	0.006	0.007	0.933
$d_e = 0.3,\ d_f = 0.3,\ d_X = 0.3$								
$N, T = 100$	-0.007	0.014	0.016	0.925	-0.006	0.016	0.017	0.928
$N, T = 200$	-0.007	0.007	0.010	0.855	-0.002	0.011	0.011	0.935
$d_e = 0.3,\ d_f = 0.4,\ d_X = 0.4$								
$N, T = 100$	-0.006	0.016	0.017	0.925	-0.006	0.019	0.020	0.925
$N, T = 200$	-0.007	0.009	0.011	0.900	-0.003	0.001	0.010	0.955

Table 2: LS estimate over frequency domain and its bias correction (DGP3 and 4)

				DGP3				
	$\hat{\beta}$ (bias-corrected time domain LS estimate by Bai (2009))				$\tilde{\beta}^{**}_{(A)}$ (bias-corrected FDLS estimate)			
	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)
				$d_e = 0,\ d_f = 0,\ d_X = 0$				
$N, T = 100$	0.000	0.004	0.004	0.955	-0.002	0.012	0.012	0.925
$N, T = 200$	0.000	0.002	0.002	0.950	-0.001	0.008	0.008	0.958
				$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$				
$N, T = 100$	-0.002	0.005	0.005	0.950	-0.005	0.014	0.015	0.940
$N, T = 200$	-0.000	0.003	0.003	0.950	-0.001	0.009	0.009	0.955
				$d_e = 0.3,\ d_f = 0.3,\ d_X = 0.3$				
$N, T = 100$	-0.010	0.008	0.013	0.755	-0.008	0.022	0.023	0.920
$N, T = 200$	-0.008	0.004	0.009	0.565	-0.004	0.015	0.015	0.948
				$d_e = 0.3,\ d_f = 0.4,\ d_X = 0.4$				
$N, T = 100$	-0.009	0.009	0.013	0.805	-0.008	0.023	0.025	0.908
$N, T = 200$	-0.008	0.005	0.010	0.595	-0.005	0.013	0.014	0.950
				DGP4				
	$\hat{\beta}$ (bias-corrected time domain LS estimate by Bai (2009))				$\tilde{\beta}^{**}_{(A)}$ (bias-corrected FDLS estimate)			
	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)
				$d_e = 0,\ d_f = 0,\ d_X = 0$				
$N, T = 100$	-0.000	0.008	0.008	0.945	-0.001	0.011	0.011	0.915
$N, T = 200$	-0.000	0.004	0.004	0.940	0.000	0.007	0.007	0.953
				$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$				
$N, T = 100$	-0.001	0.010	0.010	0.945	-0.003	0.013	0.013	0.915
$N, T = 200$	-0.001	0.005	0.005	0.945	-0.001	0.006	0.006	0.943
				$d_e = 0.3,\ d_f = 0.3,\ d_X = 0.3$				
$N, T = 100$	-0.008	0.014	0.016	0.910	-0.005	0.018	0.018	0.908
$N, T = 200$	-0.005	0.008	0.009	0.910	-0.004	0.010	0.010	0.925
				$d_e = 0.3,\ d_f = 0.4,\ d_X = 0.4$				
$N, T = 100$	-0.005	0.014	0.015	0.925	-0.004	0.017	0.018	0.923
$N, T = 200$	-0.005	0.008	0.009	0.895	-0.003	0.011	0.011	0.925

Table 3: Correction of bias

Estimator		RMSE	STD	COVP(95%)	BIAS
$\hat{\beta}$			$d_e = 0.2,\ d_f = 0,\ d_X = 0$		
	$N, T = 100$	0.001	0.001	0.965	-0.233
	$N, T = 200$	0.005	0.005	0.940	-0.381
			$d_e = 0.3,\ d_f = 0,\ d_X = 0$		
	$N, T = 100$	0.014	0.012	0.910	-0.656
	$N, T = 200$	0.008	0.006	0.845	-0.867
$\tilde{\beta}^*$			$d_e = 0.2,\ d_f = 0,\ d_X = 0$		
	$N, T = 100$	0.010	0.010	0.950	0.014
	$N, T = 200$	0.005	0.005	0.955	-0.044
			$d_e = 0.3,\ d_f = 0,\ d_X = 0$		
	$N, T = 100$	0.012	0.012	0.960	-0.037
	$N, T = 200$	0.006	0.006	0.945	0.045
$\tilde{\beta}^*_F$			$d_e = 0.2,\ d_f = 0,\ d_X = 0$		
	$N, T = 100$	0.011	0.011	0.965	0.005
	$N, T = 200$	0.005	0.005	0.965	-0.004
			$d_e = 0.3,\ d_f = 0,\ d_X = 0$		
	$N, T = 100$	0.015	0.015	0.935	-0.065
	$N, T = 200$	0.007	0.007	0.935	0.411

Table 4: Estimation of R^0

		$R^0 = 2$					$R^0 = 3$				
Method		Mean	Median	RCE	ROE	RDE	Mean	Median	RCE	ROE	RDE
IC						$d_e = 0,\ d_f = 0,\ d_X = 0$					
	$N, T = 100$	2.005	2	0.995	0.005	0	3	3	1	0	0
	$N, T = 200$	2	2	1	0	0	3	3	1	0	0
						$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$					
	$N, T = 100$	2.005	2	0.995	0.005	0	3	3	1	0	0
	$N, T = 200$	2	2	1	0	0	3	3	1	0	0
						$d_e = 0.3,\ d_f = 0.3,\ d_X = 0.3$					
	$N, T = 100$	5.255	5	0	1	0	6.225	6	0	1	0
	$N, T = 200$	4.860	5	0	1	0	5.685	6	0	1	0
						$d_e = 0.3,\ d_f = 0.4,\ d_X = 0.4$					
	$N, T = 100$	5.085	5	0	1	0	5.960	6	0	1	0
	$N, T = 200$	4.725	5	0	1	0	5.475	5	0	1	0
ER						$d_e = 0,\ d_f = 0,\ d_X = 0$					
	$N, T = 100$	1.995	2	0.995	0	0.005	3	3	1	0	0
	$N, T = 200$	2	2	1	0	0	3	3	1	0	0
						$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$					
	$N, T = 100$	1.990	2	0.990	0	0.010	3	3	1	0	0
	$N, T = 200$	2	2	1	0	0	3	3	1	0	0
						$d_e = 0.3,\ d_f = 0.3,\ d_X = 0.3$					
	$N, T = 100$	1.810	2	0.810	0	0.190	2.240	3	0.620	0	0.380
	$N, T = 200$	1.980	2	0.980	0	0.020	2.930	3	0.965	0	0.035
						$d_e = 0.3,\ d_f = 0.4,\ d_X = 0.4$					
	$N, T = 100$	1.810	2	0.810	0	0.190	2.270	3	0.635	0	0.365
	$N, T = 200$	1.980	2	0.980	0	0.020	2.910	3	0.955	0	0.045

3.7 Empirical Applications

In this section we try to adopt our methodology to the relationship between GDP and and private debt. In the literature there has been mixed results about this relationship and their long memory nature. Andrés et al. (2020) show, using an equilibrium model, that private deleveraging and slower economic growth may happen together after a tightening of fiscal policy, especially under recession. But Eggertsson and Krugman (2012) shows the negative effects from private deleveraging to economic growth could be temporary. Caporale et al. (2021) study the series of credit of non-financial sector in 43 OECD countries, and find out their highly persistent nature. Ergemen (2019) put the relationship between economic growth and private debt volume into a linear regression with interactive fixed effect like we do, but estimate the individual-specific coefficient using CCE method. In our application, we try to extend the framework of Ergemen (2019) by using a FDLS estimator and involving some other regressors, but focus on regression coefficient common to each individual.

In detail, we consider the following model:

$$\Delta \log Y_{it} = \beta_1 \Delta \log C_{it} + \beta_2 \Delta \log E_{it} + \lambda_i' F_t^0 + \varepsilon_{it}, \tag{3.7.1}$$

where λ_i and F_t^0 are factor loading and factor that are both $R \times 1$ vectors. Above we denote $\Delta \log Y_{it} = \log Y_{it} - \log Y_{i,t-1}$. And the series above could be stationary but still long memory as Caporale et al. (2021) show that most of the memory parameter estimates of debt-to-GDP ratio are close to or even slightly higher than one. The variables we involve are GDP (Y), credit to non-financial sector (C) and total employment (E) as a proxy of labor input and human capital. Our data is adopted from Bank of International Settlements (BIS) Statistics Warehouse and OECD iLibrary, including 20 OECD countries and one non-OECD country, from 1996Q1 to 2021Q1. Our panel is balanced with $T = 100$. In Table 4 we present the estimation results of β_1 to β_3 together with estimates of their standard error, as given by Theorem 3.4.3. Note that we thoroughly try $R = 1, \ldots, 8$, and as shown in Supplemental Material, the usual eigenvalue-ratio method developed by Ahn and Horenstein (2013) to estimate the number of factors still works and thus we can pick up the correct model as we wish. In Table 5 we present the descriptive statistics of our data across every country involved, where we also estimate memory parameters averaged for each variable using local-Whittle estimator within the range $\left(-\frac{1}{2}, \frac{1}{2}\right)$. We can see that for $\Delta \log Y$, almost all countries turn out to have anti-persistence; while for $\Delta \log C$ and $\Delta \log E$, some countries have these two variables to be long memory. Although so far we have not considered the setup where memory parameters are heterogeneous in regressors across individuals in our theory, the above results still partly fit our framework as long memory factors and idiosyncratic error terms can occur together with long memory regressors. Results of coefficients estimation and determination of factor number at every R (also specified as $R_{\max}$) are presented in Table 6, which shows as a long-term relationship, private debt volume has positive effect on economic growth, which also fits most of the individual effects indicated by Ergemen (2019), and eigenvalue ratio method implies there is only factor in this model.

Table 5: Descriptive statistics and memory parameter estimation in application

	$\Delta \log Y$ $\left(\hat{d} = -0.16\right)$					$\Delta \log C$ $\left(\hat{d} = 0.14\right)$					$\Delta \log E$ $\left(\hat{d} = 0.48\right)$				
Country	Mean	STD	Max	Min	Mem	Mean	STD.	Max	Min	Mem	Mean	STD	Max	Min	Mem
Spain	0.004	0.027	0.155	-0.194	-0.149	0.012	0.030	0.170	-0.115	0.276	0.004	0.012	0.033	-0.077	0.274
Belgium	0.004	0.018	0.111	-0.127	-0.272	0.012	0.022	0.116	-0.083	0.162	0.002	0.002	0.007	-0.008	0.488
Italy	0.000	0.023	0.148	-0.140	-0.233	0.007	0.025	0.171	-0.086	0.082	0.001	0.005	0.013	-0.028	0.196
Chile	0.008	0.020	0.063	-0.136	-0.022	0.016	0.026	0.068	-0.086	0.187	0.005	0.030	0.086	-0.225	-0.202
Austria	0.004	0.018	0.104	-0.122	-0.173	0.007	0.019	0.118	-0.082	-0.020	0.002	0.006	0.030	-0.043	-0.184
Hungary	0.006	0.021	0.099	-0.151	-0.075	0.014	0.038	0.132	-0.139	0.150	0.002	0.011	0.031	-0.028	-0.112
Norway	0.004	0.013	0.042	-0.047	-0.276	0.011	0.019	0.062	-0.040	0.061	0.002	0.009	0.018	-0.017	-0.160
Netherlands	0.014	0.014	0.072	-0.088	-0.067	0.008	0.018	0.073	-0.068	0.197	0.002	0.005	0.016	-0.027	0.236
Ireland	0.014	0.034	0.204	-0.054	0.017	0.023	0.064	0.493	-0.113	0.002	0.005	0.013	0.034	-0.070	0.250
France	0.003	0.024	0.171	-0.145	-0.347	0.010	0.022	0.189	-0.077	-0.391	0.002	0.004	0.014	-0.027	0.140
Sweden	0.006	0.014	0.072	-0.085	-0.052	0.014	0.018	0.083	-0.065	0.042	0.002	0.016	0.031	-0.026	-0.458
Luxembourg	0.009	0.019	0.076	-0.065	-0.017	0.023	0.045	0.207	-0.095	0.332	0.005	0.003	0.013	-0.008	0.263
Poland	0.009	0.018	0.076	-0.098	-0.280	0.022	0.032	0.105	-0.099	0.252	0.001	0.007	0.014	-0.029	0.371
Denmark	0.004	0.012	0.061	-0.067	-0.074	0.009	0.016	0.048	-0.051	0.211	0.001	0.005	0.015	-0.025	0.183
Israel	0.009	0.016	0.085	-0.097	-0.188	0.011	0.022	0.105	-0.087	0.053	0.006	0.008	0.025	-0.022	0.108
Switzerland	0.005	0.011	0.062	-0.064	-0.129	0.008	0.015	0.075	-0.035	-0.093	0.002	0.007	0.020	-0.024	-0.140
Finland	0.005	0.014	0.045	-0.067	0.105	0.010	0.019	0.065	-0.047	0.034	0.002	0.008	0.018	-0.034	0.064
Czech Republic	0.005	0.015	0.065	-0.093	0.087	0.006	0.029	0.119	-0.131	0.079	0.000	0.008	0.019	-0.021	-0.080
Portugal	0.002	0.024	0.137	-0.166	-0.216	0.008	0.026	0.148	-0.115	0.169	0.000	0.008	0.019	-0.038	0.223
United Kingdom	0.004	0.028	0.161	-0.218	-0.323	0.008	0.029	0.169	-0.163	-0.019	0.002	0.004	0.008	-0.012	0.246
Germany	0.003	0.016	0.087	-0.105	-0.227	0.003	0.016	0.102	-0.064	0.011	0.002	0.003	0.009	-0.013	0.438

Table 6: Estimation results in application

R	1	2	3	4	5	6	7	8
$\tilde{\beta}_1^{**}$	0.187	0.122	0.114	0.131	0.115	0.098	0.092	0.092
s.e.$(\tilde{\beta}_1^{**})$	(0.014)	(0.006)	(0.008)	(0.010)	(0.010)	(0.009)	(0.010)	(0.012)
$\tilde{\beta}_2^{**}$	0.167	0.162	0.161	0.122	0.086	0.151	0.115	0.123
s.e.$(\tilde{\beta}_2^{**})$	(0.026)	(0.020)	(0.031)	(0.024)	(0.025)	(0.032)	(0.029)	(0.026)
$\widehat{R}$	1	1	1	1	1	1	1	1

3.8 Conclusion

We have considered a linear regression with interactive fixed effects, where we extend the current studies as we allow for stationary long memory in regressors, factors and idiosyncratic error term. We find out the bias and convergence rate issues in classic time domain least squares estimator and its difficulty to handle in practice, and then derive our own solution to it, a frequency domain least square estimator that takes advantage of the singularity of spectral density at zero frequency for possible long memory processes. There are indeed several strands to extend the current study. First, we can discuss whether Whittle-like estimator of memory parameters of factors and idiosyncratic error term still works if using estimated factors and residual, and how would its asymptotic behavior changes. Second, as in Ergemen (2019) we can generalize our memory parameter setting to allow it to be heterogeneous in regressors across individuals, which is more relevant in application. Third we can try analyzing whether our asymptotic theories still hold under stationary anti-persistent data with $d \in \left(-\frac{1}{2}, 0\right)$, as many studies of time series long memory model have considered. Fourth it is interesting to study the long memory factor model under near-nonstationary and nonstationary data, which is more involved in application but may lead to some totally different theoretical results.

Chapter 4

Factor Memories Estimation And Test against Spurious Long Memory

4.1 Introduction

In this paper, we follow the setup in chapter two, a linear regression model with interactive fixed effects, where the regressors, the latent factors and the idiosyncratic error are allowed to be stationary long memory variables. We consider the estimation and inference of the memory parameters of the latent factors, and a test against the spurious long memory. This paper responds to the interest of panel modeling in the literature of financial and macroeconomic data (see e.g. Lahiri and Liu, 2006 and Luciani and Veredas, 2015) and the existence of long memory in such variables and some latent trends involved (see e.g. Hassler and Wolters, 1995). We adopt the traditional Local Whittle estimator developed by Robinson (1995c) to the estimated discrete Fourier transforms of the latent factor, which is based on the frequency domain principal component least squares estimator in chapter two. Estimates of factor memories are proved to be consistent and asymptotically normal without any efficiency loss relative to using the infeasible true factors. Moreover, we adopt the test developed by Qu (2011) against the existence of spurious long memory, which is characterized by short memory process contaminated by either random level shifts or time-varying smooth trends. Such processes may lead to a positive bias of the local Whittle estimator and is mistaken for a long memory process.

This paper is organized as follows. In section 4.2 we describe the data generating process and the model setup. In section 4.3 we give the local Whittle estimator of the factor memories based on the frequency domain estimate of the discrete Fourier transform of the latent factors, and analyze its asymptotic properties. In section 4.4 we establish the test against spurious long memory and derive its asymptotic null distribution. In section 4.5 we conduct the Monte Carlo simulation examining the convergence of our local Whittle estimator and the size and power under two global alternatives of the test.

The following notation is adopted. For an arbitrary $m \times n$ matrix A, its transpose is denoted by A', its conjugate is denoted by $\overline{A}$ and conjugate transpose by A^* if A is complex, its Frobenius norm is $\|A\| = \sqrt{\mathrm{tr}\,(A'A)}$ if A is real, or $\|A\| = \sqrt{\mathrm{tr}\,(A^*A)}$ if A is complex. The spectral norm of A is $\|A\|_{\mathrm{sp}} = \sqrt{\mu_1\,(A'A)}$, when A is real, and $\|A\|_{\mathrm{sp}} = \sqrt{\mu_1\,(A^*A)}$, when A is complex, where $\mu_1\,(\cdot)$ denotes the largest eigenvalue of the Hermitian matrix argument. Let $\mathbb{I}_R$ denote an R-dimensional identity matrix. For any two matrix-valued sequences A_j and B_j of the same dimension, $A_j \sim B_j$ is defined by $\frac{A_{j,(m,n)}}{B_{j,(m,n)}} \to 1$ as $j \to \infty$ for each of its (m, n)-th elements.

4.2 Model

Consider the following static panel linear regression model with interactive fixed effects,

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + \varepsilon_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T, \tag{4.2.1}$$

where we have $P \times 1$ regressors X_{it} and regression coefficients β, and $R \times 1$ latent factors F_t and factor loadings λ_i, and idiosyncratic errors ε_{it}. In this study, we allow X_{it}, F_t and ε_{it} to be all possibly stationary long memory processes with memory parameters given by $d_X = (d_{X_1}, \ldots, d_{X_P})'$, $d_F = (d_{F_1}, \ldots, d_{F_R})'$ and d_ε. In this study we focus on the scenario that the memory parameters of both X_{it} and ε_{it} are homogeneous across individuals i, so that the cross-sectional heterogeneity of memory in this model is totally explained by the interactive fixed effects, especially through multiplication of factor loadings. This setting sacrifices generality to some extent but makes it more convenient to handle the model in theory.

In terms of the data generating processes of all the possibly long memory variables, we define them as linear processes that follow Robinson and Hidalgo (1997). To be specific, let d_ε and all of the elements of d_F and d_X lie within the interval $\left[0, \frac{1}{2}\right)$. And F_t, X_{it} and ε_{it} have the following moving average representations:

$$F_t = \mu_F + \sum_{j=0}^{\infty} A_{F,j} \zeta_{F,t-j} \equiv \mu_F + F_t^o, \tag{4.2.2}$$

and for every $i = 1, \ldots, N$,

$$X_{it} = \mu_{X,i} + \sum_{j=0}^{\infty} A_{X,j} \zeta_{X,i,t-j} \equiv \mu_{X,i} + X_{it}^o, \tag{4.2.3}$$

and

$$\varepsilon_{it} = \sum_{j=0}^{\infty} A_{\varepsilon,j} \zeta_{\varepsilon,i,t-j}, \tag{4.2.4}$$

where $A_{F,j}$ and $A_{X,j}$ and are $A_{\varepsilon,j}$ respectively $R \times R$, $P \times P$ and 1×1 coefficients, and $\zeta_{F,t}$, $\zeta_{X,i,t}$ and $\zeta_{\varepsilon,i,t}$ are the corresponding innovation processes. μ_F and μ_X are $R \times 1$ and $P \times 1$ vectors of expectation. The long memory property is defined through the spectral densities of all these variables in frequency domain. For instance, let the matrix-valued spectral density function of F_t be

$f_F(\gamma)$, then it satisfies

$$f_F(\gamma) \sim \Gamma G \Gamma, \text{ as } \gamma \to 0^+,$$

where G is a positive definite matrix and $\Gamma = Diag(\gamma^{-d_{F_r}})$, $k = 1, \ldots, K$. This is the multivariate extension of the scalar spectral density of a potentially long memory process, which has the form $f(\gamma) \sim g\gamma^{-2d}$. This implies that marginally for each $k = 1, \ldots, K$, we have $f_{F_r}(\gamma) \sim g_r \gamma^{-2d_{F_r}}$. The similar power law holds for the spectral densities of X_{it} and ε_{it}.

This specification of possibly long memory processes includes the stationary fractional integrated processes and $ARFIMA\,(p, d, q)$ as special cases. And different from $\zeta_{F,t-j}$, the innovations of the regressors and the idiosyncratic error term characterize the heterogeneity and cross-sectional dependence of X_{it} and ε_{it}.

4.3 Estimation of Factor Memories

In this section, we try to estimate the memory parameters of the latent factors by local Whittle estimator using their estimated discrete Fourier transform (DFT hereafter) from frequency domain principal component least squares (FDPCLS hereafter) estimation of model (4.2.1). Then we establish the consistency and asymptotic normality of memory estimators.

4.3.1 Local Whittle estimation

To proceed, we first adopt the FDPCLS estimator of model (4.2.1), which jointly estimates the regression coefficients β and the DFT of latent factors F_t. To be specific, the FDPCLS estimator is given by the solutions of the following nonlinear equations in frequency domain:

$$\tilde{\beta} = \left[\sum_{i=1}^{N} \text{Re} \left(W_{X,i}^* M_{\hat{W}_F} W_{X,i} \right) \right]^{-1} \sum_{i=1}^{N} \text{Re} \left(W_{X,i}^* M_{\hat{W}_F} W_{Y,i} \right)$$

and

$$[W_U W_U^*] \hat{W}_F = \left[\frac{1}{NT} \sum_{i=1}^{N} \left(W_{Y,i} - W_{X,i} \tilde{\beta} \right) \left(W_{Y,i} - W_{X,i} \tilde{\beta} \right)^* \right] \hat{W}_F \quad (4.3.1)$$

$$= \hat{W}_F V_{NL}, \quad (4.3.2)$$

where V_{NL} is the diagonal matrix containing the r largest eigenvalues of $W_U W_U^*$ in decreasing order. In the above equations, $W_{X,i} = \left(W'_{X,i1}, \ldots, W'_{X,iL} \right)'$ is a $L \times P$ complex matrix that gives the DFT of regressors X_i over frequencies indexed by $j = 1, \ldots, L$. The DFT of X_i is defined by

$$W_{X,ij} = \sum_{t=1}^{T} X_{it} e^{\mathbf{i} t \gamma_j}, \; j = 1, \ldots, L,$$

where $\mathbf{i}$ is the imaginary unit such that $\mathbf{i}^2 = -1$, and $\gamma_j = \frac{2\pi j}{T}$, $j = 1, \ldots, L$ are the Fourier frequencies. Note that by conducting such DFT on both sides of model (4.2.1) with certain normalization, we can obtain a new panel linear regression model on individual and frequency dimensions, which is

$$W_{Y,ij} = \beta' W_{X,ij} + \lambda_i' W_{F,j} + W_{\varepsilon,ij}, \; i = 1, \ldots, N, \; j = 1, \ldots, L. \quad (4.3.3)$$

And thus FDPCLS is actually a frequency domain extension of the time domain least squares estimator developed by Bai (2009). To make our frequency domain estimator comparable to the time domain one, we denote $\tilde{\Gamma}_X = Diag \left\{ \gamma_L^{d_{X_p} - \frac{1}{2}} \right\}$ and $\tilde{\Gamma}_F = Diag \left\{ \gamma_L^{d_{F_r} - \frac{1}{2}} \right\}$ as part of the normalization matrix for the DFT of regressors and factors, by which we define $\tilde{W}_F = \tilde{\Gamma}_F W_F$ and $\tilde{\Lambda} = \Lambda \tilde{\Gamma}_F^{-1}$. Then similar to Bai (2009), an identifying restriction $\tilde{W}_F^* \tilde{W}_F / T = I$ is imposed to proceed with our FDPCLS estimation, and model (4.3.3) can be rewritten as

$$W_{Y,ij} = \beta' W_{X,ij} + \tilde{\lambda}_i' \tilde{W}_{F,j} + W_{\varepsilon,ij},\ i = 1, \ldots, N,\ j = 1, \ldots, L.$$

From what we have obtained above, the memories of latent factors are then estimated using the very classic local Whittle estimator because of the stationarity of the processes. To be specific, given the estimated DFT of the latent factors, $\hat{W}_F$, the local Whittle estimator minimizes the objective function marginally on each $r = 1, \ldots, R$

$$Q_r(\upsilon, d) = \frac{1}{L} \sum_{j=1}^{L} \left\{ \log\left(\upsilon \gamma_j^{-2d}\right) + \frac{\gamma_j^{2d}}{\upsilon} \hat{I}_{F_r,j} \right\}, \tag{4.3.4}$$

where $\hat{I}_{F,j} = \hat{W}_{F,j}^* \hat{W}_{F,j}$ is the periodogram of the estimated DFT of factors. Then the local Whittle estimator of $d_{F,r}$ can be written as

$$(\hat{\upsilon}_{\hat{F}_r}, \hat{d}_{\hat{F}_r}) = \arg \min_{0<\upsilon<\infty, d\in\left[0,\frac{1}{2}\right)} Q_r(\upsilon, d).$$

By first order condition in terms of υ, the estimation of d_{F_r} can also be given by

$$\hat{d}_{\hat{F}_r} = \arg \min_{d\in\left[0,\frac{1}{2}\right)} K_r\left(d\right),$$

such that for each $r = 1, \ldots, R$,

$$K_r\left(d\right) = \log \hat{G}_r\left(d\right) - 2d \frac{1}{L} \sum_{j=1}^{L} \log \gamma_j \tag{4.3.5}$$

where

$$\hat{G}_r\left(d\right) = \frac{1}{L} \sum_{j=1}^{L} \gamma_j^{2d} \hat{I}_{F_r,j}. \tag{4.3.6}$$

In the next subsection, we try to derive the asymptotic behavior of the local Whittle estimator, where we its consistency and asymptotic normality are established.

4.3.2 Asymptotic properties of the local Whittle estimator

In this subsection, we give the results of consistency and asymptotic normality of the local Whittle estimator of the factor memories, which is developed based on the FDPCLS estimator. So as in the last chapter and Bai (2009) in time domain, we define

$$\begin{aligned} D_{NL}^{\dagger}(W_F) &= \frac{1}{NT}\sum_{i=1}^{N} W_{X,i}^{*}\mathbf{M}_{W_F}W_{X,i} - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N} W_{X,i}^{*}\mathbf{M}_{W_F}W_{X,k}a_{ik} \\ &= \frac{1}{NT}\sum_{i=1}^{N} W_{Z,i}(F)^{*}W_{Z,i}(F), \end{aligned} \tag{4.3.7}$$

where $W_{Z,i}(F) = \mathbf{M}_{W_F}W_{X,i} - \frac{1}{N}\sum_{k=1}^{N}\mathbf{M}_{W_F}W_{X,k}a_{ik}$. Also let $W_{Z,i} = W_{Z,i}(F^0)$ and $D_{NL}^{\dagger} = D_{NL}^{\dagger}(W_{F^0})$. Then we can define the variable Z_{it} like we did in time domain as if its DFT over Fourier frequencies is given by $W_{Z,i}$, and denote its memory parameters by $d_Z = (d_{Z_1}, \ldots, d_{Z_P})$. Following are the assumptions to proceed with our theory:

Assumption A. (i) Denote the $(P+R+1)\times 1$ vector $V_{it} = (X_{it}', F_t', \varepsilon_{it})'$. Suppose V_{it} is covariance stationary and has the spectral density matrix satisfying

$$f_{V,i}(\gamma) \sim \Gamma(\gamma)\Upsilon_i\Gamma(\gamma) \text{ as } \gamma \to 0^+, \tag{4.3.8}$$

where Υ_i is a $(P+R+1)\times(P+R+1)$ symmetric matrix that is finite uniformly over i with the following structure:

$$\Upsilon_i = \begin{pmatrix} \Upsilon_{i,XX} & \Upsilon_{i,XF} & 0 \\ \Upsilon_{i,XF}' & \Upsilon_{FF} & 0 \\ 0 & 0 & \Upsilon_{i,\varepsilon\varepsilon} \end{pmatrix},$$

in which for all i, the $P\times P$ and $R\times R$ submatrices $\Upsilon_{i,XX}$ and $\Upsilon_{i,FF}$ are positive definite, and the scalar $\Upsilon_{i,\varepsilon\varepsilon} > 0$. Γ is a diagonal matrix given by

$$\Gamma(\gamma) = Diag\left(\gamma^{-d_{X_1}}, \ldots, \gamma^{-d_{X_P}}, \gamma^{-d_{F_1}}, \ldots, \gamma^{-d_{F_R}}, \gamma^{-d_\varepsilon}\right).$$

(ii) There exists $\theta \in (0, 2]$ such that for each i,

$$\left|f_{V,i,(ab)} - \upsilon_{i,(ab)}\gamma^{-d_a-d_b}\right| = O\left(\gamma^{\theta-d_a-d_b}\right) \text{ as } \gamma \to 0^+$$

for arbitrary $a, b = 1, \ldots, (P + R + 1)$.

(iii) Let $V_{it} = \mu_V + \sum_{j=0}^{\infty} A_{V,j}\zeta_{V,i,t-j}$, where $A_{V,j}$ is a block-diagonal matrix consisting of $A_{X,j}$, $A_{F,j}$ and $A_{\varepsilon,j}$ in order, as given by (4.2.2)-(4.2.4). And define $A_V(\gamma) = \sum_{j=0}^{\infty} A_{V,j}e^{\mathbf{i}j\gamma}$. As $\gamma \to 0^+$,

$$\left\|\frac{dA_{V,a}(\gamma)}{d\gamma}\right\| = O\left(\gamma^{-1}\left\|A_{V,a}(\gamma)\right\|\right)$$

for arbitrary $a = 1, \ldots, (P + R + 1)$, where $A_{V,a}(\gamma)$ is the a-th row of $A_{V,a}(\gamma)$.

(iv) $\zeta_{F,t}$, $\zeta_{X,t}$ and $\zeta_{\varepsilon,t}$ satisfy $E\left(\zeta_{F,t}|\mathcal{F}_{F,t-1}\right) = 0$, $E\left(\zeta_{X,t}|\mathcal{F}_{X,t-1}\right) = 0$, and $E\left(\zeta_{\varepsilon,t}|\mathcal{F}_{\varepsilon,t-1}\right) = 0$, where $\mathcal{F}_{F,t-1}$, $\mathcal{F}_{X,t-1}$ and $\mathcal{F}_{\varepsilon,t-1}$ are the corresponding filtrations.

(v) Let $\zeta_{F,t(p)}$ be the p-th element of $\zeta_{F,t}$, and the same notation applies to $\zeta_{X,t}$. We assume that $\zeta_{F,t}$ satisfy

$$E\left[\zeta_{F,t(p)}\zeta_{F,t(q)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{1,pq} < \infty,\ E\left[\zeta_{F,t(p_1)}\zeta_{F,t(p_2)}\zeta_{F,t(p_3)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{2,p_1p_2p_3} < \infty,$$

and

$$E\left[\zeta_{F,t(p_1)}\zeta_{F,t(p_2)}\zeta_{F,t(p_3)}\zeta_{F,t(p_4)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{3,p_1\ldots p_4} < \infty$$

for some absolute constants $\Phi_{1,pq}$, $\Phi_{2,p_1p_2p_3}$ and $\Phi_{3,p_1\ldots p_4}$, and for arbitrary p-, q- and p_1-, $\ldots$, p_4-th elements of $\zeta_{F,t}$. Also the same condition holds for $\zeta_{X,t}$ and

$\zeta_{\varepsilon,t}$. Additionally, $\zeta_{\varepsilon,t}$ satisfies the following eighth-order moment condition

$$E\left[\zeta_{\varepsilon,t(p_1)}\cdots\zeta_{\varepsilon,t(p_8)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{4,p_1\ldots p_8} < \infty \tag{4.3.9}$$

for some absolute constant $\Phi_{4,p_1\ldots p_8}$, and for arbitrary p_1-, …, p_8-th element of $\zeta_{\varepsilon,t}$.

(vi) $\zeta_{\varepsilon,i,t}$ is independent of $\zeta_{X,i,s}$, $\zeta_{F,r}$ and λ_j for all $r,s,t = 1,\ldots,T$ and $i,j = 1,\ldots,N$.

Assumption B. (i) Let $\check{\Gamma}_{X,j} = \mathrm{Diag}\left(\gamma_j^{d_{X_p}}\right)$, we assume $E\left\|\check{\Gamma}_{X,j}W_{X,ij}\right\|^4 \leq M$ and $\frac{1}{T}\tilde{W}_{X,i}^*\tilde{W}_{X,i} \xrightarrow{p} \Sigma_{X,i}^W > 0$ for some matrix $\Sigma_{X,i}^W$, as $T \to \infty$ for each $i = 1,\ldots,N$.

(ii) Let $\mathcal{W} = \left\{\tilde{W}_F \in \mathbb{C}^{L\times R} : \tilde{W}_F = W_F\tilde{\Gamma}_F,\ \tilde{W}_F^*\tilde{W}_F/T = \mathbb{I}_R\right\}$. We assume

$$\inf_{\tilde{W}\in\mathcal{W}} D_{NL}^{\dagger}\left(\tilde{W}\right) > 0.$$

(iii) Let $\check{\Gamma}_{F,j} = \mathrm{Diag}\left(\gamma_j^{d_{F_r}}\right)$, we assume $E\left\|\check{\Gamma}_{F,j}W_{F,j}\right\|^4 \leq M$ and $\frac{1}{T}\tilde{W}_F^*\tilde{W}_F \xrightarrow{p} \Sigma_F^W > 0$ for some matrix Σ_F^W, as $T \to \infty$.

Assumption C. (i) $E\left\|\gamma_j^{d_\varepsilon}W_{\varepsilon,ij}\right\|^8 \leq M$ and $\frac{1}{T}\tilde{W}_{\varepsilon,i}^*\tilde{W}_{\varepsilon,i} \xrightarrow{p} \Sigma_{\varepsilon,i} > 0$ for some matrix $\Sigma_{\varepsilon,i}$, as $T \to \infty$.

(ii) Let $\sqrt{E\left|W_{\varepsilon,ik}W_{\varepsilon,jl}^*\right|^2} = \sigma_{ij,kl}^{W,1}$ and $E\left|W_{\varepsilon,ik}W_{\varepsilon,il}^*\right|^2 = \sigma_{i,kl}^{W,1}$. We assume $\sigma_{ij,kl}^{W,1} \leq \gamma_k^{-d_\varepsilon}\gamma_l^{-d_\varepsilon}\overline{\sigma}_{ij}^W$ and $\sqrt{\sigma_{i,kl}^{W,1}\sigma_{j,kl}^{W,1}} \leq \gamma_k^{-2d_\varepsilon}\gamma_l^{-2d_\varepsilon}\overline{\sigma}_{ij}^W$ for all (k,l) and

$$\frac{1}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij}^W \leq M. \tag{4.3.10}$$

Moreover, let $\left|E\left(W_{\varepsilon,ik}W_{\varepsilon,jl}^*\right)\right| = \sigma_{ij,kl}^{W,2}$. We assume $\left|\sigma_{ij,kl}^{W,2}\right| \leq \overline{\sigma}_{kl}^W$ for all (i,j), and

$$\frac{\gamma_L^{2d_\varepsilon}}{L^{1+2d_\varepsilon}}\sum_{k,l=1}^{L}\overline{\sigma}_{kl}^W \leq M,\ \frac{\gamma_L^{2d_\varepsilon}}{NL^{1+2d_\varepsilon}}\sum_{i,j=1}^{N}\sum_{k,l=1}^{L}\left|\sigma_{ij,kl}^W\right| \leq M, \text{ and } \frac{1}{T^{4d_\varepsilon}\log^2 L}\sum_{k,l=1}^{L}\gamma_N^W(k,l)^2 \leq M, \tag{4.3.11}$$

where $\gamma_N^W(k,l) = \frac{1}{N}\sum_{i=1}^N E\left(W_{\varepsilon,il}W_{\varepsilon,ik}^*\right)$.

(iii) Let $\Omega_i = \check{\Gamma}_\varepsilon E\left(W_{\varepsilon,i}W_{\varepsilon,i}^*\right)\check{\Gamma}_\varepsilon$, where $\check{\Gamma}_\varepsilon = \text{Diag}\left(\gamma_j^{d_\varepsilon}\right)$. The largest eigenvalue of Ω_i is bounded uniformly over i and T as $T \to \infty$.

(iv) For every (k,l), $E\left|N^{-\frac{1}{2}}\gamma_j^{d_\varepsilon}\gamma_l^{d_\varepsilon}\sum_{i=1}^N\left[W_{\varepsilon,ik}W_{\varepsilon,il}^* - E(W_{\varepsilon,ik}W_{\varepsilon,il}^*)\right]\right|^4 \leq M$.

(v) Moreover,

$$\frac{\gamma_L^{4d_\varepsilon}}{NL^2}\sum_{i,j=1}^N\sum_{k,l=1}^L\left|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,ik}^*, W_{\varepsilon,jl}W_{\varepsilon,jl}^*)\right| \leq M, \tag{4.3.12}$$

$$\frac{\gamma_L^{4d_\varepsilon}}{N^2L^2}\sum_{i,j,m,n=1}^N\sum_{k,l=1}^L\left|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,jk}^*, W_{\varepsilon,ml}W_{\varepsilon,nl}^*)\right| \leq M,$$

and

$$\frac{\gamma_L^{4d_\varepsilon}}{NL^2}\sum_{i,j=1}^N\sum_{k,l=1}^L\left|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,il}^*, W_{\varepsilon,jk}W_{\varepsilon,jl}^*)\right| \leq M.$$

Assumption D. (i) Let $\Gamma_Z = Diag\left\{\gamma_L^{d_{Z_p}}\right\}$. $\text{plim}\,\gamma_L^{-1}\Gamma_Z D_{NL}^\dagger(W_{F^0})\Gamma_Z = D_0^W$ for some matrix $D_0^W > 0$.

(ii) $\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N W_{Z,i}^*W_{\varepsilon,i} \xrightarrow{d} \mathcal{N}(0,\Sigma)$, where

$$\Sigma^W = \text{p}\lim\frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i,j=1}^N\sum_{k,l=1}^L \text{Re}\left(\sigma_{ii,kl}^W\Gamma_Z W_{Z,ik}W_{Z,il}^*\Gamma_Z\right).$$

(iii) Moreover, we have for each $i = 1,\ldots,N$ that

$$E\left\|\frac{\sqrt{L}\gamma_L^{d_\varepsilon-\frac{1}{2}}}{T}W_{\varepsilon,i}^*\tilde{W}_{X,i}\right\|^2 \leq M, \text{ and } E\left\|\frac{\sqrt{L}\gamma_L^{d_\varepsilon-\frac{1}{2}}}{T}W_{\varepsilon,i}^*\tilde{W}_F\right\|^2 \leq M,$$

and the following also holds:

$$E\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-\frac{1}{2}}}{NT}\sum_{i=1}^N W_{\varepsilon,i}^*\tilde{W}_{X,i}\right\|^2 \leq M,$$

$$E\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-\frac{1}{2}}}{NT}\sum_{i=1}^N \lambda_i W_{\varepsilon,i}^* \tilde{W}_F\right\|^2 \leq M, \text{ and } E\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-\frac{1}{2}}}{NT}\sum_{i=1}^N W_{\varepsilon,i}^* \tilde{W}_F\right\|^2 \leq M.$$

Assumption E. As $T \to \infty$, we assume (i) $\frac{L}{T}+\frac{1}{L} \to 0$. Moreover, denote $\overline{d} = \max\{d_{X,\max}, d_{F,\max}, d_{Z,\max}\}$, $\underline{d} = \min\{d_{X,\min}, d_{F,\min}, d_{Z,\min}\}$, and $\Delta d = \overline{d} - \underline{d}$, then we assume:

(ii) $\underline{d} > d_\varepsilon$; and (iii) $\left(\sqrt{\frac{L}{N}}+\frac{1}{\sqrt{L}}\right)\gamma_L^{-2\Delta d} \to 0$.

Assumption F. The model (4.3.3) satisfies the following identifying restrictions:

(i) $\frac{1}{T}\widetilde{W}_F\widetilde{W}_F^* = I$. (ii) $\Lambda'\Lambda/N$ is a diagonal matrix with distinct main diagonal elements.

Remark 1. Assumption A-E are adopted from chapter 2, which can be treated as the frequency domain extensions of the assumptions that are imposed in Bai (2009). Assumption F corresponds to one of the identifying restrictions indicated by Bai and Ng (2013), which is also a frequency domain extension. The other two restrictions in that paper can also be applied but will not be discussed here for simplicity.

The next two theorems establish the consistency and asymptotic normality of the local Whittle estimator of each d_{F_r}, $r = 1, \ldots, R$, where from Assumption A(i), we denote υ_{F_r}, $r = 1, \ldots, R$, as the main diagonal elements of Υ_{FF}.

Theorem 4.3.1 *Suppose Assumption A-H hold, as $N, T \to \infty$ with $\frac{N}{T} \to \phi$ for some finite constant ϕ we have for each $r = 1, \ldots, R$, $\hat{d}_{\hat{F}_r} \xrightarrow{p} d_{F_r}$.*

Next we derive the asymptotic distribution of the local Whittle estimator. Denote the spectral density matrix of factor by $f_F(\gamma)$ and $\Gamma_F(\gamma) = Diag\left(\gamma^{-d_{F_1}}, \ldots, \gamma^{-d_{F_R}}\right)$, then we need the following additional assumption to proceed.

Assumption G. For some $\rho \in (0, 2]$,

$$f_F(\gamma) \sim \Gamma_F(\gamma)\Upsilon_{FF}\Gamma_F(\gamma)(1 + O(\gamma^\rho)),$$

and ρ satisfies $\frac{L^{1+2\rho}(\log L)^2}{T^{2\rho}} \to 0$.

Theorem 4.3.2 *Suppose Assumptions A-G hold, as $N, T \to \infty$ with $\frac{N}{T} \to \phi$ for some finite constant ϕ we have for each $r = 1, \ldots, R$ that $\sqrt{L}\left(\hat{d}_{\hat{F}_r} - d_{F_r}\right) \xrightarrow{d} N\left(0, \frac{1}{4}\right)$.*

4.4 Test against Spurious Long Memory

In this section we try to adopt a test from Qu (2011) against the spurious long memory of the latent factors, which is based on the score of local Whittle estimator. And then its asymptotic null distribution and consistency against several alternatives that represent the spurious long memory is presented.

We test marginally every single element of factors against spurious long memory. For each $r = 1, \ldots, R$, the null hypothesis of interest is given by

- H_0: $f_{Fr}(\gamma) \sim \upsilon_{F_r}\gamma^{-2d_{Fr}}$ as $\gamma \to 0^+$ with $\upsilon_{F_r} > 0$ and $d_{Fr} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$,

And the test statistic is based on the score of the profile likelihood function of local Whittle estimator. Recall that for each $r = 1, \ldots, R$, minimizing $Q_r(\upsilon, d)$ in (4.3.4) with respect to g helps us obtain the profiled likelihood function given by (4.3.5) and (4.3.6). The derivative of $K_r(d)$ then helps build up the test statistic, which is given by

$$\frac{\partial K_r(d_r)}{\partial d_r} = \frac{2\upsilon_r}{\sqrt{L}\hat{G}(d_r)}\left\{L^{-\frac{1}{2}}\sum_{j=1}^{L} u_j\left(\frac{\hat{I}_{F_r,j}}{\upsilon_{F_r}\gamma_j^{-2d_r}} - 1\right)\right\}$$

where $u_j = \log\gamma_j - \frac{1}{L}\sum_{j=1}^{L}\log\gamma_j = \log j - \frac{1}{L}\sum_{j=1}^{L}\log j$. Then using the above score function, the test statistic is then given by

$$V_r = \sup_{\rho\in[\iota,1]}\left(\sum_{j=1}^{L} u_j^2\right)^{-\frac{1}{2}}\left|\sum_{j=1}^{\lfloor L\rho\rfloor} u_j\left(\frac{\hat{I}_{F_r,j}}{\hat{G}\left(\hat{d}_{\hat{F}_r}\right)\gamma_j^{-2\hat{d}_{\hat{F}_r}}} - 1\right)\right|,$$

where $\hat{I}_{F,j}$ and $\hat{d}_{\hat{F}_r}$ are respectively the estimated periodogram and the local

Whittle estimator the memory parameter of the r-th latent factor from FDP-CLS; and ι is a small trimming parameter. In the following theorem, we try to establish the asymptotic null distribution of our test statistic.

Theorem 4.4.1 *Under Assumptions A-G, as $N, T \to \infty$ with $\frac{N}{T} \to \phi$ for some constant ϕ we have for each $r = 1, \ldots R$ that*

$$V_r \Rightarrow \sup_{\rho \in [\iota, 1]} \left| \int_0^\rho (1 + \log s)\, dW(s) - W(1) \int_0^\rho (1 + \log s)\, ds - \Phi(\rho) \int_0^1 (1 + \log s)\, dW(s) \right|,$$

where $\Phi(\rho) = \int_0^\rho (1 + \log s)^2\, ds$ and $W(s)$ is a Wiener process in $[0, 1]$.

4.5 Monte Carlo Simulation and Application

In this section we try to conduct an experimental simulation to testify the performance of the test above. The setup of linear regression is basically the same as in chapter 2, while under the null and alternative hypothesis, the DGPs of factors are given respectively by:

$$H_0 : F_t = 1 - 0.7F_{t-1} + e_t,$$

with e_t is an $I(d_F)$ process. And for the global alternatives, we follow the lead of Qu (2011) and consider two types of the latent factors that represent the spurious long memory. The first one is short memory processes that is contaminated by random level shifts:

$$H_1^{(1)} : F_t = \mu_t + z_t,$$

where $\mu_t = \mu_{t-1} + \pi_t \eta_t$, and π_t follows an i.i.d. Bernoulli distribution with $Pr(\pi_t = 1) = \frac{6}{T}$ and $Pr(\pi_t = 0) = 1 - \frac{6}{T}$, and $z_t \sim iidN(0, 5)$ and $\eta_t \sim iidN(0, 1)$. This alternative serves as a nonstationary random level shift, and the Bernoulli distribution ensures the amount of shift is finite even under very large sample. The second alternative is a short memory process with a time-varying smooth trend:

$$H_1^{(2)} : F_t = z_t + h\left(\frac{t}{T}\right),$$

where $h(\cdot)$ is a Lipschitz continuous function on $[0,1]$. To be specific, we specify $h\left(\frac{t}{T}\right) = \sin(4\pi t/T)$ and z_t being the same as above. In the following Table 1, we present the estimated factor memory parameter and the rate of rejections under both the null and alternative hypotheses with nominal levels of significance being 0.1 and 0.05, respectively, so as to evaluate the size and power and the performance of local Whittle estimator of the factor memory parameter.

The results are presented under three different values of d_F: 0.2, 0.3 and 0.4, with $d_X = 0.4$ and $d_\varepsilon = 0.2$ throughout these scenarios. The bandwidth L is set to be $\lfloor T^{0.7} \rfloor$. And we focus on the case where the true number of factors is 2, and in Table 1 we report the results from the second factor. Note that to implement the test, we choose the trimming parameter $\iota = 0.05$ and employ the corresponding asymptotic critical values documented in Qu (2011).

In terms of empirical application, we try to adopt the data from our last chapter and check out if there is spurious long memory through their local Whittle estimator. The results are presented for the largest estimated memory parameter of the estimated DFT of the factor, together with its test statistic. The trimming parameter is set to be 0.05 as in the simulation. We can find out that there are relatively strong evidence that support the existence of long memory in the factors, as the null is rejected under 5% level of significance under all numbers of factors, see Table 2 below. But somehow in the case $R = 1$ the estimated memory seems to fall in the range of anti-persistence, which is not so consistent with our estimation of factor number as one in last chapter. We believe a larger sample size can help fix it.

Table 1. Results of simulation

N, T	$\hat{d}_F$	5%	10%
Under H_0, $d_F = 0.2$			
$N = 100, T = 200$	0.144	0.030	0.045
$N = 200, T = 400$	0.170	0.020	0.065
$N = 300, T = 600$	0.180	0.060	0.090
Under H_0, $d_F = 0.3$			
$N = 100, T = 200$	0.239	0.030	0.040
$N = 200, T = 400$	0.259	0.045	0.065
$N = 300, T = 600$	0.275	0.060	0.085
Under H_0, $d_F = 0.4$			
$N = 100, T = 200$	0.342	0.030	0.045
$N = 200, T = 400$	0.377	0.050	0.060
$N = 300, T = 600$	0.386	0.045	0.100
Under $H_1^{(1)}$			
$N = 100, T = 200$	0.243	0.295	0.365
$N = 200, T = 400$	0.260	0.550	0.645
$N = 300, T = 600$	0.267	0.740	0.805
Under $H_1^{(2)}$			
$N = 100, T = 200$	0.191	0.425	0.565
$N = 200, T = 400$	0.200	0.725	0.805
$N = 300, T = 600$	0.209	0.950	0.980

Table 2: Results of empirical application

R	1	2	3	4	5	6	7	8
$\hat{d}_{F_{\max}}$	-0.169	0.222	0.213	0.198	0.191	0.200	0.231	0.242
$V_{\max}$	0.227	1.063	1.028	0.998	0.988	1.001	1.001	1.096

4.6 Conclusion

In this paper we prove the validity of the traditional local Whittle estimator and the test against spurious long memory for the estimated latent factor, in a long memory linear regression model with interactive fixed effects. This illustrates how the FDPCLS estimator is easy to implement not just in the estimation and inference of the regression coefficients, but also of the memory parameters.

Chapter 5

Conclusion

This dissertation contributes to the estimation and inference theory of long memory time series and panel model. We extend the current literature by considering a time-varying linear regression model and a factor model under long memory, and modify the classic frequency domain estimator to fit the new framework. The asymptotic properties of these estimators of regression coefficients and relevant tests are developed, together with Monte Carlo simulation that shows the good performance of the estimators and tests in finite sample. Also in empirical applications, our methods are employed to study the inflation spillover effect and the debt-growth nexus.

In the future research, there are several main directions to further extend this dissertation. First, a uniform inference of our time-varying regression coefficients and a data-driven bandwidth can be developed in the second chapter. Second, in the third chapter the high-level assumptions of the CLT should be replaced by the theory that generalizes the proof of Christensen and Nielsen (2006) to the factor model. Third, our analysis could be extended to the range of anti-persistence for both of our model setup, as it also covers plenty to financial data.

Appendix A

Technical Results for Chapter 2

A.1 Proofs of the Main Results

In this appendix we first state some technical lemmas and propositions whose proofs can be found later. Then we prove the main results of Chapter 2.

A.1.1 Some Technical Lemmas and Propositions

To simplify the notations, let $K_{h,tu} = K_h\left(t - Tu\right)$, $c_{Tu} = \sqrt{2\pi \sum_{t=1}^{T} K_{h,tu}^2}$, and sometimes we denote $w_{X,j} = w_X\left(u, \lambda_j\right)$ when no confusion arises from suppressing its dependence on u, and the same notation holds for $w_a\left(u, \lambda_j\right)$, $I_a\left(u, \lambda_j\right)$ and $A_a\left(u, \lambda_j\right)$ with $a = X, \varepsilon, \zeta$. First, we state three lemmas that are used in the proof of Theorem 2.3.1.

Lemma A.1.1 *Under Assumption 1-5, for any fixed $u \in (0, 1)$*

$$\Lambda_M \frac{\lambda_M^{-2\delta(u)}}{M} \sum_{j=1}^{M} Re\left(w_{X,j} w_{X,j}^*\right) \lambda_j^{2\delta(u)} \Lambda_M - \Gamma\left(u\right) = o_p\left(1\right) \tag{A.1.1}$$

where $\Lambda_M = diag\left(\lambda_M^{d_{X_1}(u)}, \ldots, \lambda_M^{d_{X_p}(u)}\right)$, and the (a, b)-th element of $\Gamma\left(u\right)$ is given by $\Gamma_{ab}\left(u\right) = G_{ab} / \left(1 - d_{X_a}\left(u\right) - d_{X_b}\left(u\right) + 2\delta\left(u\right)\right)$.

Lemma A.1.2 *Under Assumption 1-5, for any fixed $u \in (0, 1)$*

$$\Lambda_M \frac{\lambda_M^{d_\varepsilon(u) - 2\delta(u)} \sqrt{h}}{\sqrt{M}} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} Re\left(w_{X,j} w_{\varepsilon,j}^*\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega\left(u\right)\right)$$

where $\Lambda_M = diag\left(\lambda_M^{d_{X_1}(u)}, \ldots, \lambda_M^{d_{X_p}(u)}\right)$, *and the* (a,b)*-th element of* $\Omega(u)$ *is*

$$\Omega_{ab}(u) = \Theta^*\left(\frac{G_{X,aa}^{\frac{1}{2}} G_{X,bb}^{\frac{1}{2}} G_{\varepsilon\varepsilon}}{2\left(1 - d_{X_a}(u) - d_{X_b}(u) - 2d_\varepsilon(u) + 4\delta(u)\right)}\right),$$

given a finite deterministic function $\Theta^*(\cdot)$.

The following two positions are used in the proof of Theorem 2.3.2:

Proposition A.1.3 *Under Assumption 1 to 5, for any* M *that is defined in Assumption 4(i), we have*

(i) $S_a(m) = \frac{1}{T}\sum_{j=1}^{M} \left|w_{X_a,j} - A_{X_a,j} w_{\zeta,j}\right|^2 = O_p\left(hF_{X_{aa}}(u, \lambda_m)\right)$,

(ii) $S_b(m) = \frac{1}{T}\sum_{j=1}^{M}\left(I_{X_{aa},j} + A_{X_a,j} I_{\zeta,j} A_{X_a,j}^*\right) = O_p\left(hF_{X_{aa}}(u, \lambda_m)\right)$,

where $w_{X_a,j}$ *is the a-th element of* $w_{X,j}$*, and the same notation holds for* $A_{X_a,j}$ *and* $I_{X_{aa},j}$*; and* $F_{X_{aa}}(u, \lambda_m)$ *is the marginal spectral distribution function corresponding to the spectral density defined in Assumption 2.*

Proposition A.1.4 *Let*

$$\widetilde{B}_{1,Ma} = \sum_{j=1}^{M} \lambda_j^{2\delta(u)} Re\left(I_{a\varepsilon,j} - A_{X_a,j} I_{\zeta,j} A_{\varepsilon,j}^*\right),$$

where $I_{a\varepsilon,j}$ *is the cross-periodogram between* X_a *and* ε*. Then suppose Assumption 1-5 hold, we have for each* $a = 1, \ldots, p$,

$$\widetilde{B}_{1,Ma} = O_p\left(\lambda_M^{2\delta(u) - d_{X_a}(u) - d_\varepsilon(u)}\left(h^{-1}(\log M)^2 + h^{-1}M^{\frac{1}{4}}(\log M)^{\frac{1}{2}} + h^{-1}M^{\frac{1}{2}}T^{-\frac{1}{4}}(\log M)^{\frac{1}{2}}\right)\right).$$

Proposition A.1.5 *Define*

$$Z_{t,T}(u) = \frac{\zeta_t' K_{h,tu}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} K_{h,tu}^2}} \sum_{s<t} C_{t-s,T}(u) \frac{\zeta_s K_{h,su}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} K_{h,tu}^2}},$$

where $C_{t-s,T}(u)$ *is given by*

$$C_{t-s,T}(u) = \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u) + d_\varepsilon(u) - 2\delta(u)}\sqrt{h}}{2\pi T\sqrt{M}} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} Re\left(A_{X_a,j}' \overline{A}_{\varepsilon,j} + A_{\varepsilon,j}' \overline{A}_{X_a,j}\right)\cos(t\lambda_j)$$

with arbitrary weight η_a*. Given a finite deterministic function* $\Theta^*(\cdot)$*, recall*

$$\Omega_{ab} = \Theta^*\left(\frac{G_{X,aa}^{\frac{1}{2}} G_{X,bb}^{\frac{1}{2}} G_\varepsilon}{1 - d_{X_a}(u) - d_{X_b}(u) - 2d_\varepsilon(u) + 4\delta(u)}\right),$$

then under the conditions of Theorem 2.3.1, the following conclusions will hold:

(i) $\sum_{t=1}^{T}\mathbb{E}\left(Z_{t,T}^{2}(u)\mid\mathcal{F}_{t-1}\right)\xrightarrow{p}\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\Omega_{ab}$;

(ii) $\sum_{t=1}^{T}\mathbb{E}\left(Z_{t,T}^{2}(u)\mathbf{1}\left(|Z_{t,T}(u)|>\varepsilon\right)\right)\xrightarrow{p}0$ *for all* $\varepsilon>0$.

A.1.2 Proofs of the Main Theorems

Proof of Theorem 2.3.1. Recall that

$$\begin{aligned}\hat{\beta}(u) &= \left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,w_X^*(u,\lambda_j)\right)\right]^{-1}\left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,w_y^*(u,\lambda_j)\right)\right]\\ &\equiv R_{XX}^{-1}(u)\,R_{XY}(u)\,.\end{aligned}$$

Using $y_t=\beta(u)'X_{t,T}+\left(\beta\left(\frac{t}{T}\right)-\beta(u)\right)'X_{t,T}+\varepsilon_{t,T}$, we make the following decomposition for $w_y(u,\lambda_j)$:

$$\begin{aligned}w_y(u,\lambda_j) &= \beta(u)'\frac{1}{c_{Tu}}\sum_{t=1}^{T}X_{t,T}K_{h,tu}e^{it\lambda_j}+\frac{1}{c_{Tu}}\sum_{t=1}^{T}\left(\beta\left(\frac{t}{T}\right)-\beta(u)\right)'X_{t,T}K_{h,tu}e^{it\lambda_j}\\ &\quad+\frac{1}{c_{Tu}}\sum_{t=1}^{T}\varepsilon_{t,T}K_{h,tu}e^{it\lambda_j}\\ &\equiv \beta'(u)\,w_X(u,\lambda_j)+\tilde{w}_X(u,\lambda_j)+w_\varepsilon(u,\lambda_j)\,,\end{aligned}$$

then it follows that

$$\begin{aligned}\hat{\beta}(u)-\beta(u) &= R_{XX}^{-1}(u)\left\{\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,\tilde{w}_X^*(u,\lambda_j)\right)+\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,w_\varepsilon^*(u,\lambda_j)\right)\right\}\\ &\equiv R_{XX}^{-1}(u)\left\{\mathbb{B}_M(u)+\mathbb{V}_M(u)\right\},\end{aligned}$$

where $\mathbb{B}_M(u)$ and $\mathbb{V}_M(u)$ contribute to the asymptotic bias and variance parts of $\hat{\beta}(u)-\beta(u)$ respectively.

First for $\mathbb{V}_M(u)$, by Lemma A.1.1 we have

$$\Lambda_M\frac{\lambda_M^{-2\delta(u)}}{M}R_{XX}(u)\,\Lambda_M=\Lambda_M\frac{\lambda_M^{-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,w_X^*(u,\lambda_j)\right)\Lambda_M\xrightarrow{p}\Gamma(u)\,. \tag{A.1.2}$$

By Lemma A.1.2,

$$\Lambda_M\frac{\lambda_M^{d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,w_\varepsilon^*(u,\lambda_j)\right)\xrightarrow{d}N\left(0,\Omega(u)\right).$$

Next, we study $\mathbb{B}_M(u)$. By Assumption 3(i), $\left|\frac{t}{T}-u\right|\leq h$ holds uniformly on the support of kernel function $k(\cdot)$. Then by Assumption 5 and Taylor expansion, we have that for any

$\left|\frac{t}{T}-u\right|=O\left(h\right),$

$$\beta\left(\frac{t}{T}\right)=\beta\left(u\right)+\beta^{(1)}\left(u\right)\left(\frac{t}{T}-u\right)+O\left(h^{2}\right),$$

where $\beta^{(1)}\left(u\right)$ is the first order derivative of $\beta\left(u\right)$. Note that to derive the order of bias term, it is sufficient to replace $\beta\left(\frac{t}{T}\right)-\beta\left(u\right)$ with $\beta^{(1)}\left(u\right)\left(\frac{t}{T}-u\right)$ as it is the leading term. So in the following we denote instead

$$\tilde{w}_{X}\left(u,\lambda_{j}\right)=\frac{1}{c_{Tu}}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)X_{t,T}K_{h,tu}e^{it\lambda_{j}},$$

and

$$\mathbb{B}_{M}\left(u\right)=\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\mathrm{Re}\left(w_{X}\left(u,\lambda_{j}\right)\tilde{w}_{X}^{*}\left(u,\lambda_{j}\right)\right)\beta^{(1)}\left(u\right).$$

Uniformly in $u\in[0,1]$, we define and decompose $\mathbb{B}_{M}\left(u\right)$ as

$$\begin{aligned}\mathbb{B}_{M}\left(u\right)=&\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(A_{X}\left(u,\lambda_{j}\right)w_{\zeta}\left(u,\lambda_{j}\right)\widetilde{w}_{\zeta}^{*}\left(u,\lambda_{j}\right)A_{X}^{*}\left(u,\lambda_{j}\right)\right)\beta^{(1)}\left(u\right)\\&+\left(\mathbb{B}_{M}\left(u\right)-\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(A_{X}\left(u,\lambda_{j}\right)w_{\zeta}\left(u,\lambda_{j}\right)\widetilde{w}_{\zeta}^{*}\left(u,\lambda_{j}\right)A_{X}^{*}\left(u,\lambda_{j}\right)\right)\beta^{(1)}\left(u\right)\right)\\&\equiv\left(B_{1}+B_{2}\right)\beta^{(1)}\left(u\right),\end{aligned}$$

where we define $\widetilde{w}_{\zeta}\left(u,\lambda_{j}\right)=\frac{1}{c_{Tu}}\sum_{t=1}^{T}\zeta_{t}\left(\frac{t}{T}-u\right)K_{h,tu}e^{it\lambda_{j}}$. We analyze B_1 and B_2 in the following. Firstly for B_1 we have

$$\begin{aligned}B_{1}=&\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(A_{X}\left(u,\lambda_{j}\right)w_{\zeta}\left(u,\lambda_{j}\right)\widetilde{w}_{\zeta}^{*}\left(u,\lambda_{j}\right)A_{X}^{*}\left(u,\lambda_{j}\right)\right)\\=&\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(A_{X}\left(u,\lambda_{j}\right)\frac{1}{c_{Tu}^{2}}\sum_{t,s=1}^{T}\zeta_{t}\zeta_{s}^{'}\left(\frac{s}{T}-u\right)K_{h,tu}K_{h,su}e^{i(t-s)\lambda_{j}}A_{X}^{*}\left(u,\lambda_{j}\right)\right)\\=&\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(A_{X}\left(u,\lambda_{j}\right)\frac{1}{c_{Tu}^{2}}\sum_{t=1}^{T}\zeta_{t}\zeta_{t}^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^{2}A_{X}^{*}\left(u,\lambda_{j}\right)\right)\\&+\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(A_{X}\left(u,\lambda_{j}\right)\frac{1}{c_{Tu}^{2}}\sum_{t\neq s}\zeta_{t}\zeta_{s}^{'}\left(\frac{s}{T}-u\right)K_{h,tu}K_{h,su}e^{i(t-s)\lambda_{j}}A_{X}^{*}\left(u,\lambda_{j}\right)\right)\\\equiv&\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(b_{1,j}\right)+\sum_{j=1}^{M}\lambda_{j}^{2\delta(u)}\,\mathrm{Re}\left(b_{2,j}\right)\equiv B_{11}+B_{12},\end{aligned}$$

then we can closely follow the proof of Lemma A.1.1 to derive their order. To be specific, firstly we still consider an arbitrary (a,b)-th element of B_{11}, denoted as $B_{11,ab}$, whose expectation is given by

$$\mathbb{E}\left(B_{11,ab}\right)=\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\operatorname{Re}\left(A_{X_a}\left(u,\lambda_j\right)A_{X_b}^{*}\left(u,\lambda_j\right)\right)\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2$$
$$\leq h\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\operatorname{Re}\left(f_{X_{ab}}\left(u,\lambda_j\right)\right)=O\left(hM\lambda_M^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\right),$$

and its variance is given by

$$Var\left(B_{11,ab}\right)=\sum_{j,k=1}^{M}\lambda_j^{2\delta(u)}\lambda_k^{2\delta(u)}Cov\left(\operatorname{Re}\left(b_{1,j,ab}\right),\operatorname{Re}\left(b_{1,k,ab}\right)\right)$$
$$\leq\sum_{j,k=1}^{M}\lambda_j^{2\delta(u)}\lambda_k^{2\delta(u)}\left\{Var\left(\operatorname{Re}\left(b_{1,j,ab}\right)\right)Var\left(\operatorname{Re}\left(b_{1,k,ab}\right)\right)\right\}^{\frac{1}{2}},$$

where it is adequate to consider $Var\left(\operatorname{Re}\left(b_{1,j}\right)\right)$ only. Then we have

$$Var\left(\operatorname{Re}\left(b_{1,j,ab}\right)\right)$$
$$=\mathbb{E}\left(\operatorname{Re}\left(b_{1,j,ab}\right)-\mathbb{E}\left(\operatorname{Re}\left(b_{1,j,ab}\right)\right)\right)^2$$
$$\leq\mathbb{E}\left|A_{X_a}\left(u,\lambda_j\right)\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\left(\zeta_t\zeta_t^{'}-I_{p+1}\right)\left(\frac{t}{T}-u\right)K_{h,tu}^2A_{X_b}^{*}\left(u,\lambda_j\right)\right|^2$$
$$\equiv\mathbb{E}\left|A_{X_a,j}\widetilde{D}_{1,j}A_{X_b,j}^{*}\right|^2,$$

where I_{p+1} is a $(p+1)\times(p+1)$ identity matrix, then

$$\mathbb{E}\left|A_{X_a,j}\widetilde{D}_{1,j}A_{X_b,j}^{*}\right|^2=\mathbb{E}\left(A_{X_a,j}\widetilde{D}_{1,j}A_{X_b,j}^{*}A_{X_b,j}\widetilde{D}_{1,j}A_{X_a,j}^{*}\right)$$
$$=\mathbb{E}\left[tr\left(A_{X_a,j}\widetilde{D}_{1,j}A_{X_b,j}^{*}A_{X_b,j}\widetilde{D}_{1,j}A_{X_a,j}^{*}\right)\right]$$
$$=vec\left(A_{X_a,j}^{*}A_{X_a,j}\right)^{'}\mathbb{E}\left(\widetilde{D}_{1,j}\otimes\widetilde{D}_{1,j}\right)vec\left(A_{X_b,j}^{*}A_{X_b,j}\right),$$

where $\mathbb{E}\left(\widetilde{D}_{1,j}\otimes\widetilde{D}_{1,j}\right)$ can be further given by

$$\mathbb{E}\left(\widetilde{D}_{1,j}\otimes\widetilde{D}_{1,j}\right) \tag{A.1.3}$$
$$=\mathbb{E}\left[\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\left(\zeta_t\zeta_t^{'}-I_p\right)\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\left(\zeta_t\zeta_t^{'}-I_p\right)\left(\frac{t}{T}-u\right)K_{h,tu}^2\right]$$
$$=\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\right)$$
$$-\frac{I_{p+1}}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\right)$$
$$-\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\right)\otimes\frac{I_{p+1}}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2$$

$$+\frac{I_{p+1}}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\frac{I_{p+1}}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2.$$

Note that $\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2=O\left(h\right)$ by domain of the kernel function, and

$$\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\right)=\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2=O\left(h\right)I_{p+1},$$

thus

$$\begin{aligned}&\mathbb{E}\left(\widetilde{D}_{1,j}\otimes\widetilde{D}_{1,j}\right)\\&=\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\right)\\&-\frac{I_{p+1}}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\frac{I_{p+1}}{c_{Tu}^2}\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}^2,\end{aligned}$$

and then

$$\begin{aligned}&\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\right)\\&=\frac{1}{c_{Tu}^4}\sum_{t,s=1}^{T}\mathbb{E}\left(\zeta_t\zeta_t^{'}\otimes\zeta_s\zeta_s^{'}\right)\left(\frac{t}{T}-u\right)\left(\frac{s}{T}-u\right)K_{h,tu}^2K_{h,su}^2.\end{aligned}$$

Compared with (A.2.4), the above equation has extra components $\left(\frac{t}{T}-u\right)$ and $\left(\frac{s}{T}-u\right)$ that are uniformly bounded by h. Then given the non-negativity of $K_{h,tu}^2K_{h,su}^2$ and the fact that $\mathbb{E}\left(\zeta_t\zeta_t^{'}\otimes\zeta_s\zeta_s^{'}\right)$ has a uniform bound as well, we can adjust the results in (A.2.4) and conclude that

$$\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}\left(\frac{t}{T}-u\right)K_{h,tu}^2\right)=O\left(\frac{h}{T}\right)\left(I_{p+1}\otimes I_{p+1}\right)$$

and so as the other terms in (A.1.3). Therefore $\mathbb{E}\left(\widetilde{D}_{1,j}\otimes\widetilde{D}_{1,j}\right)=O\left(\frac{h}{T}\right)\left(I_{p+1}\otimes I_{p+1}\right)$, and therefore

$$\begin{aligned}Var\left(\mathrm{Re}\left(b_{1,j,ab}\right)\right)&\leq\mathbb{E}\left|A_{X_a,j}\widetilde{D}_{1,j}A_{X_b,j}^{*}\right|^2\\&=vec\left(A_{X_a,j}^{*}A_{X_a,j}\right)^{'}\mathbb{E}\left(\widetilde{D}_{1,j}\otimes\widetilde{D}_{1,j}\right)vec\left(A_{X_b,j}^{*}A_{X_b,j}\right)\\&=O\left(\frac{h}{T}\lambda_j^{-d_{X_a}(u)-d_{X_b}(u)}\lambda_k^{-d_{X_a}(u)-d_{X_b}(u)}\right)\end{aligned}$$

and

$$
\begin{aligned}
Var\left(B_{11,ab}\right) &\leq \sum_{j,k=1}^{M} \lambda_j^{2\delta(u)} \lambda_k^{2\delta(u)} \left\{Var\left(\text{Re}\left(b_{1,j,ab}\right)\right) Var\left(\text{Re}\left(b_{1,k,ab}\right)\right)\right\}^{\frac{1}{2}} \\
&= O\left(\frac{h}{T} M^2 \lambda_M^{2\left[2\delta(u)-d_{X_a}(u)-d_{X_b}(u)\right]}\right).
\end{aligned}
$$

Combining the order of $\mathbb{E}\left(B_{11,ab}\right)$ and $Var\left(B_{11,ab}\right)$, we can conclude that $B_{11} = O_p\left(hM\lambda_M^{2\delta(u)-2\overline{d}_X(u)}\right)$ with $\overline{d}_X\left(u\right) = \max_p d_{X_p}\left(u\right)$.

Next for B_{12}, we can also closely follow the corresponding parts in the proof of Lemma A.1.1. We denote

$$
\begin{aligned}
B_{12} &= \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \text{Re}\left(A_X\left(u,\lambda_j\right) \frac{1}{c_{Tu}^2} \sum_{t\neq s} \zeta_t \zeta_s^{'} \left(\frac{s}{T}-u\right) K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j} A_X^*\left(u,\lambda_j\right)\right) \\
&= \frac{1}{c_{Tu}^2} \sum_{t\neq s} \zeta_t^{'} \text{Re}\left(\sum_{j=1}^{M} \lambda_j^{2\delta(u)} A_X^{'}\left(u,\lambda_j\right) \overline{A}_X\left(u,\lambda_j\right) \left(\frac{s}{T}-u\right) K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j}\right) \zeta_s \\
&\equiv \frac{1}{c_{Tu}^2} \sum_{t\neq s} \zeta_t^{'} \text{Re}\left(\widetilde{\Phi}_{t,s,M}\right) \zeta_s.
\end{aligned}
$$

Since $\mathbb{E}\left(B_{12}\right) = 0$, we only need to consider its variance, which is bounded by

$$
\begin{aligned}
&\frac{1}{c_{Tu}^4} \mathbb{E}\left|\sum_{t\neq s} \zeta_t^{'} \text{Re}\left(\widetilde{\Phi}_{t,s,M}\right) \zeta_s\right|^2 \\
&= \frac{1}{c_{Tu}^4} \sum_{t_1\neq s_1} \sum_{t_2\neq s_2} \text{vec}\left(\overline{\widetilde{\Phi}}_{t_2,s_2,M}\right)^{'} \mathbb{E}\left(\zeta_{s_2}\zeta_{s_1}^{'} \otimes \zeta_{t_2}\zeta_{t_1}^{'}\right) \text{vec}\left(\widetilde{\Phi}_{t_1,s_1,M}\right).
\end{aligned}
$$

The above equation has the same structure as (A.2.5), then the reasoning for analyzing V_1 and V_2 in the proof of Lemma A.1.1 still holds. And we can easily see that the only difference lies within the order of two following objects:

$$
\left|\frac{1}{c_{Tu}^2} \sum_{t=1}^{T} \left(\frac{t}{T}-u\right)^2 K_{h,tu}^2 e^{it\lambda_{j-k}}\right| = o\left(h^{-1}\theta\left(u,j\right)\right)
$$

and

$$
\left|\frac{1}{c_{Tu}^2} \sum_{t=1}^{T} \left(\frac{t}{T}-u\right) K_{h,tu}^2 e^{it\lambda_{j-k}}\right| = o\left(h^{-1}\theta\left(u,j\right)\right)
$$

with $j \neq k$. These two orders hold by Assumption 3(iv), and they correspond to $\left|\frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 e^{it\lambda_{j-k}}\right|$ in V_1 and V_2 respectively. The rest of the reasoning will follow the proof of Lemma A.1.1 and we can conclude that $B_{12} = o_p\left(\sqrt{M} h^{-\frac{1}{2}} \lambda_M^{2\delta(u)-2\overline{d}_X(u)}\right)$.

Next recall that B_2 is given by

$$B_2 = \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \operatorname{Re}\left(w_X(u,\lambda_j)\,\tilde{w}_X^*(u,\lambda_j) - A_X(u,\lambda_j)\,w_\zeta(u,\lambda_j)\,\widetilde{w}_\zeta^*(u,\lambda_j)\,A_X^*(u,\lambda_j)\right)$$
$$\equiv \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \operatorname{Re}(b_{3,j}).$$

Note that the above equation has the same structure as $Q_{1,M}$ in (A.2.1) in the proof of Lemma A.1.1, whose order is derived using Proposition A.1.3. Then following the same reasoning, we have for B_2 that

$$|B_2| \le \sum_{j=1}^{M-1}\left(\lambda_j^{2\delta(u)} - \lambda_{j+1}^{2\delta(u)}\right)\sum_{k=1}^{M} |\operatorname{Re}(b_{3,k})| + \lambda_M^{2\delta(u)} \sum_{j=1}^{M} |\operatorname{Re}(b_{3,j})|,$$

where considering an arbitrary (a,b)-th element and using a more compact notation,

$$\sum_{j=1}^{M} |\operatorname{Re}(b_{3,j,ab})|$$
$$\le \sum_{j=1}^{M} \left| w_{X_a,j}\tilde{w}_{X_b,j}^* - A_{X_a,j} w_{\zeta,j}\widetilde{w}_{\zeta,j}^* A_{X_b,j}^*\right|$$
$$= \sum_{j=1}^{M} \frac{1}{2}\left|(w_{X_a,j} - A_{X_a,j}w_{\zeta,j})\left(\tilde{w}_{X_b,j}^* + \widetilde{w}_{\zeta,j}^* A_{X_b,j}^*\right) + (w_{X_a,j} + A_{X_a,j}w_{\zeta,j})\left(\tilde{w}_{X_b,j}^* - \widetilde{w}_{\zeta,j}^* A_{X_b,j}^*\right)\right|,$$

then following how we analyze the order of $S_1(M)$ and $S_2(M)$ in the proof of Lemma A.1.1, it is sufficient to analyze the square root of the following two objects:

$$\widetilde{B}_{21} = \sum_{j=1}^{M} |w_{X_a,j} - A_{X_a,j}w_{\zeta,j}|^2 \text{ and } \widetilde{B}_{22} = \sum_{j=1}^{M} |\tilde{w}_{X_b,j} - \widetilde{w}_{\zeta,j}A_{X_b,j}|^2,$$

where the first term can directly follow Proposition A.1.3 and thus $\widetilde{B}_{21} = o_p(TF_{X_{aa}}(u,\lambda_M))$. Then for $\widetilde{B}_{22}$, we can still closely follow the proof of Proposition A.1.3, and only difference appears in the counterparts of (A.2.10), which is given by

$$K_*(u,\lambda) = \frac{1}{c_{Tu}^2}\left|\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}e^{it\lambda}\right|^2,$$

where

$$\left|\sum_{t=1}^{T}\left(\frac{t}{T}-u\right)K_{h,tu}e^{it\lambda}\right|$$
$$= \left|\sum_{t=1}^{T-1}\left[\left(\frac{t}{T}-u\right)K_{h,tu} - \left(\frac{t+1}{T}-u\right)K_{h,t+1,u}\right]\sum_{s=1}^{t} e^{is\lambda} + (1-u)\,K_{h,T,u}\sum_{t=1}^{T} e^{it\lambda}\right|$$

$$\leq \sum_{t=1}^{T-1} \left| \left(\frac{t}{T} - u \right) K_{h,tu} - \left(\frac{t+1}{T} - u \right) K_{h,t+1,u} \right| \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} e^{is\lambda} \right| + (1-u) K_{h,T,u} \left| \sum_{t=1}^{T} e^{it\lambda} \right|$$

$$= \sum_{t=1}^{T-1} \left| \left(\frac{t}{T} - u \right) (K_{h,tu} - K_{h,t+1,u}) - \frac{1}{T} K_{h,t+1,u} \right| \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} e^{is\lambda} \right| + (1-u) K_{h,T,u} \left| \sum_{t=1}^{T} e^{it\lambda} \right|$$

$$\leq \left(h \sum_{t=1}^{T-1} |K_{h,tu} - K_{h,t+1,u}| + \frac{1}{T} \sum_{t=1}^{T-1} K_{h,t+1,u} \right) \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} e^{is\lambda} \right| + h K_{h,T,u} \left| \sum_{t=1}^{T} e^{it\lambda} \right|.$$

Since for some constant C, $\frac{1}{T} \sum_{t=1}^{T-1} K_{h,t+1,u} \leq C$ by Assumption 3(i); and by the proof of Proposition A.1.3, $\sum_{t=1}^{T-1} |K_{h,tu} - K_{h,t+1,u}| \leq C h^{-1}$ and $\left| \sum_{s=1}^{t} e^{is\lambda} \right| \leq \frac{C}{|\lambda|}$ uniformly over t, we can thus imply that

$$\left| \sum_{t=1}^{T} \left(\frac{t}{T} - u \right) K_{h,tu} e^{it\lambda} \right| \leq \frac{C}{|\lambda|},$$

and thus $K_* (u, \lambda) = O \left(\frac{h}{T} |\lambda|^{-2} \right)$, which is the order of $\widetilde{K} (u, \lambda)$ multiplying h^2 in (A.2.11). Note that this order holds uniformly over j, so we can directly follow the proof of Proposition A.1.3 and conclude that $\widetilde{B}_{22} = o_p \left(T h^2 F_{X_{bb}} (u, \lambda_M) \right)$, and then following how we analyze (A.2.3) in the proof of Lemma A.1.1,

$$\sum_{j=1}^{M} |\text{Re} (b_{3,j,ab})| = o_p \left(T h F_{X_{aa}}^{\frac{1}{2}} (u, \lambda_M) F_{X_{bb}}^{\frac{1}{2}} (u, \lambda_M) \right)$$

and thus

$$B_2 = \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \text{Re} (b_{3,j}) = o_p \left(h M \lambda_M^{2\delta(u) - 2\overline{d}_X(u)} \right).$$

Finally, the whole bias term can be given by

$$\begin{aligned} R_{XX}^{-1} (u) \mathbb{B}_M (u) &= \left(\frac{\lambda_M^{2\overline{d}_X(u) - 2\delta(u)}}{M} R_{XX} (u) \right)^{-1} \left(\frac{\lambda_M^{2\overline{d}_X(u) - 2\delta(u)}}{M} \mathbb{B}_M (u) \right) \\ &= O_p (1) \left(O_p (h) + o_p \left(\frac{1}{\sqrt{Mh}} \right) + o_p (h) \right) \beta^{(1)} (u) \\ &= (O_p (h) + o_p (h)) \beta^{(1)} (u) = O_p (h) \beta^{(1)} (u), \end{aligned}$$

where the second equality holds by (A.1.2) and the third one holds by Assumption 4(iii). This result determines the order of bias term under our local-constant-type estimator. And note that our proof above also indicates this order to hold uniformly over $u \in (0, 1)$, which will help our proof of Theorem 2.3.2 in the following. Therefore denoting $\widetilde{\mathbb{B}}_M (u) = R_{XX}^{-1} (u) \mathbb{B}_M (u)$, we can conclude that

$$\Lambda_M^{-1} (u) \lambda_M^{d_\varepsilon(u)} \sqrt{Mh} \left(\hat{\beta} (u) - \beta (u) - h \widetilde{\mathbb{B}}_M (u) \right) \xrightarrow{d} \mathcal{N} \left(0, \Gamma^{-1} (u) \Omega (u) \Gamma^{-1} (u) \right),$$

where

$$\widetilde{\mathbb{B}}_M(u) = h^{-1}\left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,w_X^*(u,\lambda_j)\right)\right]^{-1} \times\left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,\tilde{w}_X^*(u,\lambda_j)\right)\right]\beta^{(1)}(u).$$

Summarizing the above results completes the proof of Theorem 2.3.1. ■

Proof of Theorem 2.3.2. Without loss of generality, we assume $p = 1$ in this proof for notation simplicity. Recall that in this case $\hat{\beta}(u)$ is given by

$$\hat{\beta}(u) = \left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,\overline{w}_X(u,\lambda_j)\right)\right]^{-1}\left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,\overline{w}_y(u,\lambda_j)\right)\right] \equiv \frac{R_{XY}(u)}{R_{XX}(u)}.$$

Then following the proof of Theorem 2 and 8 in Hansen (2008) we set $\widetilde{R}^0_{XY}(u) = \beta(u)\,\Gamma(u)$ with $\Gamma(u)$ defined as in Lemma A.1.1 and thus $\Gamma(u) = \frac{G_X}{1-2d_X(u)+2\delta(u)}$ with G_X being the constant multiplier of the spectral density of $X_{t,T}$, as we define in (2.3.4) in Assumption 2(i). Then we can further denote

$$\hat{\beta}(u) = \frac{R_{XY}(u)}{R_{XX}(u)} = \frac{\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}R_{XY}(u)}{\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}R_{XX}(u)} \equiv \frac{\widetilde{R}_{XY}(u)}{\widetilde{R}_{XX}(u)} = \frac{\widetilde{R}_{XY}(u)/\Gamma(u)}{\widetilde{R}_{XX}(u)/\Gamma(u)},$$

where $d_X(u)$ is the memory parameter local to a fixed $u \in [0,1]$ for the scalar regressor $X_{t,T}$. Note that the denominator is further given by

$$\frac{\widetilde{R}_{XX}(u)}{\Gamma(u)} = 1 + \frac{\widetilde{R}_{XX}(u)-\Gamma(u)}{\Gamma(u)};$$

and the numerator is further given by

$$\frac{\widetilde{R}_{XY}(u)}{\Gamma(u)} = \beta(u) + \frac{\widetilde{R}_{XY}(u)-\widetilde{R}^0_{XY}(u)}{\Gamma(u)}.$$

Therefore in the following it is sufficient to prove the uniform asymptotic negligibility of $\widetilde{R}_{XX}(u)-\Gamma(u)$ and derive the uniform order of $\widetilde{R}_{XY}(u)-\widetilde{R}^0_{XY}(u)$ over $u \in \mathcal{U}$.

First, we denote $a_{MT} = \sqrt{\frac{\log M}{Th}}$, then we try to show that $\sup_{u\in\mathcal{U}}\left|\widetilde{R}_{XX}(u)-\Gamma(u)\right| =$

$O_p(a_{MT})$. Denote

$$A_X(u) = \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(A_{X,j} I_{\zeta,j} A_{X,j}^*\right)$$

using the same notations as in the proof before. Note that

$$\sup_{u\in\mathcal{U}} \left|\widetilde{R}_{XX}(u) - \Gamma(u)\right| \leq \sup_{u\in\mathcal{U}} \left|\widetilde{R}_{XX}(u) - A_X(u)\right| + \sup_{u\in\mathcal{U}} \left|A_X(u) - \mathbb{E}(A_X(u))\right| + \sup_{u\in\mathcal{U}} \left|\mathbb{E}(A_X(u)) - \Gamma(u)\right|,$$

and by the proof of Lemma A.1.1, $\sup_{u\in\mathcal{U}} \left|\widetilde{R}_{XX}(u) - A_X(u)\right|$ and $\sup_{u\in\mathcal{U}} \left|\mathbb{E}(A_X(u)) - \Gamma(u)\right|$ correspond to $Q_{1,M}$ and $Q_{3,M}$ in (A.2.1). Therefore we have

$$\widetilde{R}_{XX}(u) = A_X(u) + O_p(h) \text{ and } \mathbb{E}(A_X(u)) = \Gamma(u) + o(h)$$

uniformly over $u \in \mathcal{U}$. Then it remains to study the order of $\sup_{u\in\mathcal{U}} \left|A_X(u) - \mathbb{E}(A_X(u))\right|$, which is given by

$$\sup_{u\in\mathcal{U}} \left|A_X(u) - \mathbb{E}(A_X(u))\right| = \sup_{u\in\mathcal{U}} \left| \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(A_{X,j}\left(I_{\zeta,j} - \mathbb{E}(I_{\zeta,j})\right) A_{X,j}^*\right) \right|$$

where for each $j = 1, \ldots, M$

$$\begin{aligned}
\mathrm{Re}\left(A_{X,j} I_{\zeta,j} A_{X,j}^*\right) &= \mathrm{Re}\left(A_{X,j} \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \zeta_t \zeta_s^{'} K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j} A_{X,j}^*\right) \\
&= A_{X,j} \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} \zeta_t \zeta_t^{'} K_{h,tu}^2 A_{X,j}^* + \mathrm{Re}\left(A_{X,j} \frac{1}{c_{Tu}^2} \sum_{t\neq s} \zeta_t \zeta_s^{'} K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j} A_{X,j}^*\right) \\
&= A_{X,j} \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} \zeta_t \zeta_t^{'} K_{h,tu}^2 A_{X,j}^* + \mathrm{Re}\left(A_{X,j} \frac{1}{c_{Tu}^2} 2 \sum_{t>s} \zeta_t \zeta_s^{'} K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j} A_{X,j}^*\right) \\
&\equiv \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} \left(K_{h,tu}^2 Z_{1t,j} + 2K_{h,tu} Z_{2t,j}\right).
\end{aligned}$$

It is easy to see for each j, the process $\{Z_{1t,j}\}$ is i.i.d. over t with its expectation equal to $A_{X,j} A_{X,j}^*$. And

$$\begin{aligned}
\mathrm{Re}\left(A_{X,j} \mathbb{E}(I_{\zeta,j}) A_{X,j}^*\right) &= \mathrm{Re}\left(A_{X,j} \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}\left(\zeta_t \zeta_s^{'}\right) K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j} A_{X,j}^*\right) \\
&= A_{X,j} \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 A_{X,j}^* = \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \mathbb{E}(Z_{1t,j}).
\end{aligned}$$

And $Z_{2t,j}$ is given by

$$Z_{2t,j} = \zeta_t^{'} \sum_{s=1}^{t-1} K_{h,su} \mathrm{Re} \left(A_{X,j}^{'} \overline{A}_{X,j} e^{i(t-s)\lambda_j} \right) \zeta_s,$$

which implies that $\{Z_{2t,j}\}$ is a martingale difference sequence (m.d.s.). Therefore $\sup_{u\in\mathcal{U}} |A_X(u) - \mathbb{E}(A_X(u))|$ can be bounded as the following,

$$\begin{aligned}
\sup_{u\in\mathcal{U}} |A_X(u) - \mathbb{E}(A_X(u))| &\leq \sup_{u\in\mathcal{U}} \left| \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re} \left(\frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 (Z_{1t,j} - \mathbb{E}(Z_{1t,j})) \right) \right| \\
&+ \sup_{u\in\mathcal{U}} \left| \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re} \left(\frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu} Z_{2t,j} \right) \right| \\
&= \sup_{u\in\mathcal{U}} \left| \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \left[\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}(Z_{1t,j} - \mathbb{E}(Z_{1t,j})) \right] \right| \\
&+ \sup_{u\in\mathcal{U}} \left| \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu} \left[\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}(Z_{2t,j}) \right] \right| \\
&\equiv \sup_{u\in\mathcal{U}} \left| \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \ddot{Z}_{1t,M} \right| + \sup_{u\in\mathcal{U}} \left| \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu} \ddot{Z}_{2t,M} \right| \\
&\equiv \sup_{u\in\mathcal{U}} |SX_1(u)| + \sup_{u\in\mathcal{U}} |SX_2(u)| .
\end{aligned}$$

Therefore using the restriction of linear process given by Assumption 1(i), we can return to analyzing the objects with average over time domain, and thus can borrow some classic methodology to obtain its uniform order (e.g. Hansen (2008)). In the following we firstly study $\sup_{u\in\mathcal{U}} |SX_1(u)|$. As the standard procedures to derive its uniform convergence rate, we firstly truncate the process $\left\{ \ddot{Z}_{1t,M} \right\}$ using the threshold $\tau_{MT} = a_{MT}^{-1}$. Then we denote for any $u \in \mathcal{U}$,

$$\begin{aligned}
SX_1(u) &= \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \ddot{Z}_{1t,M} \mathbf{1} \left(\left| \ddot{Z}_{1t,M} \right| \leq \tau_{MT} \right) + \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \ddot{Z}_{1t,M} \mathbf{1} \left(\left| \ddot{Z}_{1t,M} \right| > \tau_{MT} \right) \\
&\equiv SX_{11}(u) + SX_{12}(u) .
\end{aligned}$$

For $SX_{12}(u)$ we have

$$\begin{aligned}
\mathbb{E} |SX_{12}(u)| &\leq \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \mathbb{E} \left(\left| \ddot{Z}_{1t,M} \right| \mathbf{1} \left(\left| \ddot{Z}_{1t,M} \right| > \tau_{MT} \right) \right) \\
&\leq \frac{1}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \mathbb{E} \left(\left| \ddot{Z}_{1t,M} \right|^2 \tau_{MT}^{-(2-1)} \mathbf{1} \left(\left| \ddot{Z}_{1t,M} \right| > \tau_{MT} \right) \right) \\
&\leq \frac{\tau_{MT}^{-1}}{c_{Tu}^2} \sum_{t=1}^{T} K_{h,tu}^2 \mathbb{E} \left| \ddot{Z}_{1t,M} \right|^2 \leq \tau_{MT}^{-1},
\end{aligned}$$

where in the second inequality we adopt Assumption 6(ii) with $s=2$, and the last inequality holds by the fact that

$$\begin{aligned}\mathbb{E}\left|\ddot{Z}_{1t,M}\right|^{s} &= \mathbb{E}\left|\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(Z_{1t,j}-\mathbb{E}\left(Z_{1t,j}\right)\right)\right|^{s} \\ &= \mathbb{E}\left|\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X,j}\left(\zeta_t\zeta_t^{'}-I_2\right)A_{X,j}^{*}\right)\right|^{s} \\ &\leq \left|\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\left|A_{X,j}\right|^2\right|^{s}\mathbb{E}\left\|\zeta_t\zeta_t^{'}-I_2\right\|^{s}\end{aligned}$$

using Assumption 2(i) and 6, and definition of c_{Tu}^2. Then Markov's inequality implies that $|SX_{12}(u)| = O_p\left(\tau_{MT}^{-1}\right) = O_p\left(a_{MT}\right)$ holds uniformly over $u\in\mathcal{U}$.

Then in the following for $SX_{11}(u)$ we can assume $\left|\ddot{Z}_{1t,M}\right|\leq\tau_T$ and replace $\ddot{Z}_{1t,M}\mathbf{1}\left(\left|\ddot{Z}_{1t,M}\right|\leq\tau_{MT}\right)$ with $\ddot{Z}_{1t,M}$. To further proceed, we split the space $\mathcal{U}$ using regions $U_i=\left\{u:u\in\mathcal{U},\ |u-u_i|\leq a_{MT}h^2\right\}$ with grids $\{u_i\}$, $i=1,\ldots,N$. Then by definition $N=\frac{1}{a_{MT}h^2}$ asymptotically. Suppose we define $\overline{K}_{h,tu}=K_{h,tu}^2$, then the nonnegativity, boundedness and Lipschitz continuity of $K_{h,tu}$ from Assumption 3(i) and 3(ii) together ensure the Lipschitz continuity of $\overline{K}_{h,tu}$. Thus we can treat $\overline{K}_{h,tu}\equiv h^{-2}\overline{k}\left(\frac{t-Tu}{Th}\right)$ as a new kernel. Note that as in the proof of Proposition A.1.3, Hansen (2008, pp. 740-741) implies that for all $u\in U_i$ and any i, the kernel $\overline{K}_{h,tu}$ satisfies that

$$\left|\overline{K}_{h,tu}-\overline{K}_{h,tu_i}\right|\leq Ch^{-2}\left|u-u_i\right|\overline{K}_{h,tu_i}^{*}\leq Ca_{MT}\overline{K}_{h,tu_i}^{*}$$

for some constant C and some integrable kernel function $\overline{K}_{h,tu}^{*}$. And for $SX_{11}(u)$, we also have

$$\begin{aligned}\left|SX_{11}(u)\right| &= \left|\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}K_{h,tu}^2\ddot{Z}_{1t,M}\right| \qquad\qquad (\text{A.1.4}) \\ &= \left|\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\overline{K}_{h,tu}\left[\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X,j}\left(\zeta_t\zeta_t^{'}-I_2\right)A_{X,j}^{*}\right)\right]\right| \\ &\leq \left|\frac{C}{2\pi Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu}\left[\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\left|A_{X,j}\right|^2\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right]\right| \\ &= \left|\frac{C}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu}\left[\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}f_X\left(u,\lambda_j\right)\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right]\right| \\ &= \left|\frac{C}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu}\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right|\left(1+O_p\left(\lambda_M^{\gamma}\right)\right),\end{aligned}$$

where the first inequality holds by the definition of c_{Tu}^2 and Riemann sum approximation, and the last two equalities hold by Assumption 2(i) and 2(ii). Then for any i,

$$\begin{aligned}\sup_{u\in U_i}|SX_{11}(u)| &= \sup_{u\in U_i}\left|\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}K_{h,tu}^2\ddot{Z}_{1t,M}\right| \\ &\leq \sup_{u\in U_i}\left|\frac{C}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu}\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right|(1+O_p(\lambda_M^{\gamma})) \\ &\equiv \sup_{u\in U_i}|\Psi(u)|(1+O_p(\lambda_M^{\gamma})),\end{aligned}$$

with $\gamma\geq 1$ defined in Assumption 2(ii). So to derive the order of $\sup_{u\in U_i}|SX_{11}(u)|$ for each i and then for $\sup_{u\in\mathcal{U}}|SX_{11}(u)|$, it is sufficient to study $\sup_{u\in U_i}|\Psi(u)|$ for each i and then for $\sup_{u\in\mathcal{U}}|\Psi(u)|$. Then $\sup_{u\in U_i}|\Psi(u)|$ can be further given by

$$\begin{aligned}\sup_{u\in U_i}|\Psi(u)| &= \sup_{u\in U_i}\left|\frac{C}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu}\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right| \\ &\leq \left|\frac{C}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu_i}\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right| + Ch^{-2}|u-u_i|\left|\frac{C}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu_i}^{*}\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right|,\end{aligned}$$

thus

$$\begin{aligned}\Pr\left(\sup_{u\in\mathcal{U}}|\Psi(u)|\geq\varepsilon\right) &\leq N\max_{1\leq i\leq N}\Pr\left(\sup_{u\in U_i}|\Psi(u)|\geq\varepsilon\right) \\ &\leq N\max_{1\leq i\leq N}\Pr\left(\left|\frac{1}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu_i}\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right|\geq C\varepsilon\right) \qquad \text{(A.1.5)} \\ &+ N\max_{1\leq i\leq N}\Pr\left(\left|\frac{1}{Th^{-1}}\sum_{t=1}^{T}\overline{K}_{h,tu_i}^{*}\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right|\geq C\varepsilon\right), \qquad \text{(A.1.6)}\end{aligned}$$

where in (A.1.6) we use the fact that $a_{MT}<1$ when T is large enough. In the following we bound both (A.1.5) and (A.1.6) using the same exponential inequality adopted from Bennett (1962, pp. 34) as

$$\Pr\left(\left|\sum_{t=1}^{T}\overline{k}\left(\frac{t-Tu}{Th}\right)\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right|\geq CTh\varepsilon\right)\leq\exp\left(-\frac{C^2\varepsilon^2/2}{\sigma^2+Q_TC\varepsilon/3}\right),$$

where we take (A.1.5) for instance and use $\overline{K}_{h,tu}=h^{-2}\overline{k}\left(\frac{t-Tu}{Th}\right)$, and the above inequality holds for every $\varepsilon\geq 0$ with $\sigma^2=\sum_{t=1}^{T}\mathbb{E}\left(\overline{k}^2\left(\frac{t-Tu}{Th}\right)\left\|\zeta_t\zeta_t^{'}-I_2\right\|^2\right)$, and $Q_T=\max_{1\leq t\leq T}Q_t$ with $\overline{k}\left(\frac{t-Tu}{Th}\right)\left\|\zeta_t\zeta_t^{'}-I_2\right\|\leq Q_t$. Then to proceed, we can set $Q_T=\tau_{MT}=a_{MT}^{-1}$ using the same reasoning as in (A.1.4). And a bound for σ^2 can hold with some constant C as

$$\sigma^2=\sum_{t=1}^{T}\mathbb{E}\left(\overline{k}^2\left(\frac{t-Tu}{Th}\right)\left\|\zeta_t\zeta_t^{'}-I_2\right\|^2\right)\leq CTh$$

by Riemann sum approximation. Then by setting $\varepsilon = a_{MT} = \sqrt{\frac{\log M}{Th}}$, we can obtain for sufficiently large T that

$$Pr\left(\left|\sum_{t=1}^{T}\overline{k}\left(\frac{t-Tu}{Th}\right)\left\|\zeta_t\zeta_t^{'}-I_2\right\|\right|\geq CTha_{MT}\right)\leq \exp\left(-\frac{C^2T^2h^2a_{MT}^2/2}{Th+C^2a_{MT}^{-1}Tha_{MT}/3}\right)$$
$$=\exp\left(-\frac{\left(C^2\log M\right)/2}{1+C^2/3}\right)=M^{-\frac{C^2}{2+\frac{2}{3}C^2}},$$

which holds uniformly over u. Then

$$\Pr\left(\sup_{u\in\mathcal{U}}|\Psi(u)|\geq\varepsilon\right)\leq NM^{-\frac{C^2}{2+\frac{2}{3}C^2}}=a_{MT}^{-1}h^{-2}M^{-\frac{C^2}{2+\frac{2}{3}C^2}}=\left(\frac{T}{(\log M)\,M^{\frac{C^2}{1+\frac{1}{3}C^2}}h^3}\right)^{\frac{1}{2}}=o(1)$$

when C is sufficiently large by Assumption 7. Therefore we can conclude that $\sup_{u\in\mathcal{U}}|SX_1(u)| = O_p(a_{MT})$.

Next for $\sup_{u\in\mathcal{U}}|SX_2(u)|$, we have by what we obtained from the proof of Lemma A.1.2,

$$\begin{aligned}
&\sup_{u\in\mathcal{U}}|SX_2(u)|\\
&=\sup_{u\in\mathcal{U}}\left|\frac{1}{c_{Tu}^2}\sum_{t=2}^{T}K_{h,tu}\ddot{Z}_{2t,M}\right|\\
&=\sup_{u\in\mathcal{U}}\left|\frac{1}{c_{Tu}^2}\sum_{t=2}^{T}K_{h,tu}\left[\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(Z_{2t,j}\right)\right]\right|\\
&=\sup_{u\in\mathcal{U}}\left|\frac{1}{c_{Tu}^2}\sum_{t=2}^{T}K_{h,tu}\zeta_t^{'}\sum_{s<t}K_{h,su}\frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\,\mathrm{Re}\left(2A_{X,j}^{'}\overline{A}_{X,j}\right)\cos\left((t-s)\lambda_j\right)\zeta_s\right|\\
&\equiv\sup_{u\in\mathcal{U}}\left|\sum_{t=2}^{T}\widetilde{Z}_{2t,M}\right|.
\end{aligned}$$

Note that $\widetilde{Z}_{2t,M}$ is a m.d.s., then we can analyze its order using the same truncation threshold, and the same split of $\mathcal{U}$, while using a different exponential inequality for martingale (see e.g. (1.9) in Fan et al. (2015)), which is given by

$$P\left\{S_T\geq\varepsilon\quad and\quad V_T\leq v^2\quad for\quad some\quad T\right\}\leq\exp\left(-\frac{\varepsilon^2}{2\left(v^2+Q_T\varepsilon\right)}\right)$$

if the m.d.s. satisfies that $\max_{1\leq t\leq T}|X_t|\leq Q_T$ with probability approaching one, with $S_T=\sum_{t=1}^{T}X_t$ and $V_T=\sum_{t=1}^{T}\mathbb{E}\left(X_t^2\mid\mathcal{F}_{t-1}\right)$. Then under our setup

$$\sum_{t=1}^{T}\mathbb{E}\left(\widetilde{Z}_{2t,M}^2\mid\mathcal{F}_{t-1}\right)$$

$$= \sum_{t=1}^{T} \mathbb{E}\left(\left(\frac{K_{h,tu}}{c_{Tu}^2}\zeta_t' \sum_{s<t} K_{h,su} \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \operatorname{Re}\left(2A_{X,j}'\overline{A}_{X,j}\right) \cos\left(\left(t-s\right)\lambda_j\right)\zeta_s\right)^2 \mid \mathcal{F}_{t-1}\right)$$

$$= \sum_{t=1}^{T} \breve{Z}_{2t}'\breve{Z}_{2t}$$

where

$$\begin{aligned}
\breve{Z}_{2t} &= \frac{K_{h,tu}}{c_{Tu}^2} \sum_{s<t} K_{h,su} \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \operatorname{Re}\left(2A_{X,j}'\overline{A}_{X,j}\right) \cos\left(\left(t-s\right)\lambda_j\right)\zeta_s \\
&\equiv \frac{K_{h,tu}}{c_{Tu}^2} \sum_{s<t} \Upsilon_{Ms}\zeta_s.
\end{aligned}$$

Then by the nonnegativity of $\sum_{t=1}^{T} \breve{Z}_{2t}'\breve{Z}_{2t}$, we have

$$\begin{aligned}
\mathbb{E}\left|\sum_{t=1}^{T} \breve{Z}_{2t}'\breve{Z}_{2t}\right| &= \sum_{t=1}^{T} \mathbb{E}\left(\breve{Z}_{2t}'\breve{Z}_{2t}\right) \\
&= \sum_{t=1}^{T} \frac{K_{h,tu}^2}{c_{Tu}^4} \mathbb{E}\left[tr\left(\sum_{s<t}\sum_{r<t} \zeta_r'\Upsilon_{Mr}'\Upsilon_{Ms}\zeta_s\right)\right] = \sum_{t=1}^{T} \frac{K_{h,tu}}{c_{Tu}^2} \sum_{s<t} \left\|\Upsilon_{Ms}\right\|^2 \\
&\leq \sum_{t=1}^{T} \frac{K_{h,tu}^2}{c_{Tu}^4} \sum_{s<t} \left(K_{h,su} \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \left\|\operatorname{Re}\left(2A_{X,j}'\overline{A}_{X,j}\right) \cos\left(\left(t-s\right)\lambda_j\right)\right\|\right)^2 \\
&\leq \sum_{t=1}^{T} \frac{K_{h,tu}^2}{c_{Tu}^4} \sum_{s<t} \left(K_{h,su} \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)-2d_X(u)} \cos\left(\left(t-s\right)\lambda_j\right)\right)^2 \\
&\leq C
\end{aligned}$$

for some constant C, where the second last inequality follows from how we derived the asymptotic covariance in the proof of Proposition A.1.4. Then by Markov's inequality

$$\Pr\left(\sum_{t=1}^{T} \breve{Z}_{2t}'\breve{Z}_{2t} > \varphi\right) \leq \frac{C}{\varphi}$$

holds for any $\varphi \geq 0$, which means $\sum_{t=1}^{T} \breve{Z}_{2t}'\breve{Z}_{2t}$ is bounded in probability. Therefore using the above exponential inequality we have for any $u \in \mathcal{U}$ that

$$\begin{aligned}
\Pr\left(\left|\sum_{t=1}^{T} \widetilde{Z}_{2t,M}\right| \geq \varepsilon\right) &= P\left\{\left|\sum_{t=1}^{T} \ddot{Z}_{2t,M}\right| \geq \varepsilon \text{ and } \sum_{t=1}^{T} \mathbb{E}\left(\ddot{Z}_{2t}^2\right) \leq C \text{ for some T}\right\} \\
&\leq \exp\left(-\frac{\varepsilon^2}{2\left(C+Q_T\varepsilon\right)}\right) = M^{-\frac{C^2}{2+\frac{2}{3}C^2}}
\end{aligned}$$

by specifying $\varepsilon = \varepsilon_T = \sqrt{\frac{\log M}{Th}}$, which is the same order as we obtained just now. This helps concludes that $\sup_{u\in\mathcal{U}} \left|SX_2\left(u\right)\right| = O_p\left(a_{MT}\right)$. Therefore we have $\sup_{u\in\mathcal{U}} \left|\widetilde{R}_{XX}\left(u\right) - \Gamma\left(u\right)\right| =$

$O_p(a_{MT}) + o_p(1)$.

Second, we try to derive the uniform order of $\sup_{u\in\mathcal{U}}\left|\widetilde{R}_{XY}(u) - \widetilde{R}^0_{XY}(u)\right|$, which can be further given by

$$\sup_{u\in\mathcal{U}}\left|\widetilde{R}_{XY}(u) - \widetilde{R}^0_{XY}(u)\right| \leq \sup_{u\in\mathcal{U}}\left|\widetilde{R}_{XX}(u)\beta(u) - \widetilde{R}^0_{XY}(u)\right| + \sup_{u\in\mathcal{U}}\left|\overline{\widetilde{R}}_{XX}(u)\right| + \sup_{u\in\mathcal{U}}\left|\widetilde{R}_{X\varepsilon}(u)\right|,$$

where

$$\overline{\widetilde{R}}_{XX}(u) = \frac{\lambda_M^{2d_X(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,\tilde{w}_X^*(u,\lambda_j)\right)$$

with $\tilde{w}_X(u,\lambda_j)$ defined as part of bias term in the proof of Theorem 2.3.1. Since

$$\sup_{u\in\mathcal{U}}\left|\widetilde{R}_{XX}(u)\beta(u) - \widetilde{R}^0_{XY}(u)\right| = \sup_{u\in\mathcal{U}}\left|\widetilde{R}_{XX}(u)\beta(u) - \Gamma(u)\beta(u)\right| = O_p(a_{MT}) + o_p(1)$$

and $\sup_{u\in\mathcal{U}}\left|\overline{\widetilde{R}}_{XX}(u)\right| = O_p(h)$. Then it remains to study the order of $\sup_{u\in\mathcal{U}}\left|\widetilde{R}_{X\varepsilon}(u)\right|$. Using the same reasoning and by replacing $w_X(u,\lambda_j)$ with $w_\varepsilon(u,\lambda_j)$, we can obtain

$$\sup_{u\in\mathcal{U}}\left|\widetilde{R}_{X\varepsilon}(u)\right| = O_p\left(\lambda_M^{\triangle d}a_{MT}\right).$$

Therefore by gathering all the terms we have so far, $\sup_{u\in\mathcal{U}}\left|\hat{\beta}(u) - \beta(u)\right| = O_p(h) + O_p\left(\lambda_M^{\triangle d}\sqrt{\frac{\log M}{Th}}\right)$. ■

Proof of Theorem 2.4.1. Following the reasoning and notations in the proof of Theorem 2.3.1, we firstly decompose $\widehat{\beta}_*(u)$, which is

$$\begin{aligned}\widehat{\beta}_*(u) &= \left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\,\mathrm{Re}\left(w_X(u,\lambda_j)\,w_X^*(u,\lambda_j)\right)\right]^{-1}\left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\,\mathrm{Re}\left(w_X(u,\lambda_j)\,w_{y_*}^*(u,\lambda_j)\right)\right]\\ &\equiv R_{XX}^{-1}(u)\,R_{XY_*}(u).\end{aligned}$$

And then within $R_{XY_*}(u)$, $w_{y_*}^*(u,\lambda_j)$ is given by

$$\begin{aligned}w_{y_*}^*(u,\lambda_j) &= \frac{1}{c_{Tu}}\sum_{t=1}^{T}K_{h,tu}y_{*t,T}e^{it\lambda_j}\\ &= \frac{1}{c_{Tu}}\sum_{t=1}^{T}K_{h,tu}\left(X_{t,T}^{'}\hat{\beta}\left(\frac{t}{T}\right) + \varepsilon_{*t,T}\right)e^{it\lambda_j}\\ &= \frac{1}{c_{Tu}}\sum_{t=1}^{T}K_{h,tu}\left[X_{t,T}^{'}\left(\hat{\beta}(u) + \left(\hat{\beta}\left(\frac{t}{T}\right) - \hat{\beta}(u)\right)\right) + \varepsilon_{*t,T}\right]e^{it\lambda_j}\\ &\equiv w_X^*(u,\lambda_j)\,\hat{\beta}(u) + w_{\varepsilon_*}^*(u,\lambda_j) + \widehat{w}_X^*(u,\lambda_j),\end{aligned}$$

where $w^*_{\varepsilon_*}(u,\lambda_j)$ is defined in the same manner as $w^*_{y_*}(u,\lambda_j)$, and $\widehat{w}^*_X(u,\lambda_j)$ is defined by

$$\begin{aligned}\widehat{w}^*_X(u,\lambda_j) &= \frac{1}{c_{Tu}}\sum_{t=1}^{T} K_{h,tu}X'_{t,T}\left(\hat{\beta}\left(\frac{t}{T}\right)-\hat{\beta}(u)\right)e^{it\lambda_j} \\ &= \frac{1}{c_{Tu}}\sum_{t=1}^{T} K_{h,tu}X'_{t,T}\left[\left(\hat{\beta}\left(\frac{t}{T}\right)-\beta\left(\frac{t}{T}\right)\right)-\left(\hat{\beta}(u)-\beta(u)\right)\right]e^{it\lambda_j} \\ &- \frac{1}{c_{Tu}}\sum_{t=1}^{T} K_{h,tu}X'_{t,T}\left(\beta\left(\frac{t}{T}\right)-\beta(u)\right)e^{it\lambda_j} \\ &\equiv \widehat{w}^{*(1)}_X(u,\lambda_j)+\widetilde{w}^*_X(u,\lambda_j),\end{aligned}$$

where $\widetilde{w}^*_X(u,\lambda_j)$ is identical to the one defined in the proof of Theorem 2.3.1 about the order of bias. And by Theorem 2.3.2, both $\hat{\beta}\left(\frac{t}{T}\right)-\beta\left(\frac{t}{T}\right)$ and $\hat{\beta}(u)-\beta(u)$ have uniform order given by $O_p(h)$. Thus we can follow the same reasoning as in the proof of Theorem 2.3.1 to obtain the order of bias as $O_p(h)$. In detail,

$$\hat{\beta}^*(u) = \hat{\beta}(u) + h\widetilde{\mathbb{B}}^*_M(u) + R^{-1}_{XX}(u)\,R_{X\varepsilon_*}(u)$$

where

$$\widetilde{\mathbb{B}}_{*M}(u) = h^{-1}\left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,w^*_X(u,\lambda_j)\right)\right]^{-1}\left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(w_X(u,\lambda_j)\,\widehat{w}^*_X(u,\lambda_j)\right)\right] = O_p(1)$$

and $R_{X\varepsilon_*}(u) = \left[\sum_{j=1}^{M}\lambda_j^{2\delta(u)}Re\left(w_X(u,\lambda_j)\,w^*_{\varepsilon_*}(u,\lambda_j)\right)\right]$. Note that the asymptotics of $R^{-1}_{XX}(u)$ is derived as in Lemma A.1.1, then it remains to analyze the limiting distribution of $R_{X\varepsilon_*}(u)$, which follows that

$$\begin{aligned}&R_{X\varepsilon_*}(u)\\ &= \Lambda_M(u)\,\lambda_M^{d_\varepsilon(u)-2\delta(u)}\sqrt{\frac{h}{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\,\mathrm{Re}\left(w_X(u,\lambda_j)\,w^*_{\varepsilon_*}(u,\lambda_j)\right)\\ &= \Lambda_M(u)\,\lambda_M^{d_\varepsilon(u)-2\delta(u)}\sqrt{\frac{h}{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\,\mathrm{Re}\left(A_X(u,\lambda_j)\,I_{\zeta_*}(u,\lambda_j)\,A^*_\varepsilon(u,\lambda_j)\right) \qquad \text{(A.1.7)}\\ &+ \Lambda_M(u)\,\lambda_M^{d_\varepsilon(u)-2\delta(u)}\sqrt{\frac{h}{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\,\mathrm{Re}\left(w_X(u,\lambda_j)\,w^*_{\varepsilon_*}(u,\lambda_j)-A_X(u,\lambda_j)\,I_{\zeta_*}(u,\lambda_j)\,A^*_\varepsilon(u,\lambda_j)\right)\\ &\qquad \text{(A.1.8)}\end{aligned}$$

together with the convergence rate. Since (A.1.7), conditional on $(\boldsymbol{X},\boldsymbol{\varepsilon})$, can follow the same moment condition and Linderberg condition as in the proof of Lemma A.1.2, as we only replace the innovation vector process $\{\zeta_t\}$ with the pseudo innovation vector process $\{\zeta_{*t}\}$

that share the same moment conditions up to fourth order. Then it remains to analyze the asymptotics of (A.1.8). Following we have done in the proof of Lemma A.1.2 and Proposition A.1.4, an arbitrary a-th element of (A.1.8) can be further given by

$$\begin{aligned}
&\sum_{j=1}^{M} \lambda_j^{2\delta(u)} \operatorname{Re}\left(w_{X_a}(u,\lambda_j) w_{\varepsilon_*}^*(u,\lambda_j) - A_{X_a}(u,\lambda_j) I_{\zeta_*}(u,\lambda_j) A_{\varepsilon}^*(u,\lambda_j)\right) \\
&= \sum_{j=1}^{M-1} \left(\lambda_j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)} - \lambda_{j+1}^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\right) \sum_{k=1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \operatorname{Re}\left(I_{a\varepsilon_*,k} - A_{X_a,k} I_{\zeta_*,k} A_{\varepsilon,k}^*\right) \\
&+ \lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)} \sum_{j=1}^{M} \lambda_j^{d_{X_a}(u)+d_\varepsilon(u)} \operatorname{Re}\left(I_{a\varepsilon_*,j} - A_{X_a,j} I_{\zeta_*,j} A_{\varepsilon,j}^*\right) \\
&\equiv \sum_{j=1}^{M-1} \left(\lambda_j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)} - \lambda_{j+1}^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\right) Q^*(j) + \lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)} Q^*(M)
\end{aligned}$$

using more compact notation as before, and here we do not consider the multiplier $\Lambda_M(u)\lambda_M^{d_\varepsilon(u)-2\delta(u)}\sqrt{\frac{h}{M}}$ for now. Then as in the proof of Proposition A.1.4, $Q^*(j)$ for any fixed $1 \le j \le M$ follows that

$$\begin{aligned}
Q^*(j) &= \sum_{k=1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \operatorname{Re}\left(I_{a\varepsilon_*,k} - A_{X_a,k} I_{\zeta_*,k} A_{\varepsilon,k}^*\right) \\
&= \sum_{k=1}^{l} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \operatorname{Re}\left(I_{a\varepsilon_*,k} - A_{X_a,k} I_{\zeta_*,k} A_{\varepsilon,k}^*\right) + \sum_{k=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \operatorname{Re}\left(I_{a\varepsilon_*,k} - A_{X_a,k} I_{\zeta_*,k} A_{\varepsilon,k}^*\right) \\
&\equiv Q_1^* + Q_2^*.
\end{aligned}$$

Firstly for Q_1^* we have

$$\begin{aligned}
\mathbb{E}\left|Q_1^*\right| \le \frac{1}{2}&\left(\sum_{k=1}^{l} \lambda_k^{d_{X_a}(u)} \mathbb{E}\left|w_{X_a,k} - A_{X_a,k} w_{\zeta_*,k}\right|^2\right)^{\frac{1}{2}} \frac{1}{2}\left(\sum_{k=1}^{l} \lambda_k^{d_\varepsilon(u)} \mathbb{E}\left|w_{\varepsilon_*,k}^* + w_{\zeta_*,k}^* A_{\varepsilon,k}^*\right|^2\right)^{\frac{1}{2}} \\
+ \frac{1}{2}&\left(\sum_{k=1}^{l} \lambda_k^{d_{X_a}(u)} \mathbb{E}\left|w_{X_a,k} + A_{X_a,k} w_{\zeta_*,k}\right|^2\right)^{\frac{1}{2}} \frac{1}{2}\left(\sum_{k=1}^{l} \lambda_k^{d_\varepsilon(u)} \mathbb{E}\left|w_{\varepsilon_*,k}^* - w_{\zeta_*,k}^* A_{\varepsilon,k}^*\right|^2\right)^{\frac{1}{2}}.
\end{aligned}$$

At each $1 \le k \le l$,

$$\begin{aligned}
&\mathbb{E}\left|w_{X_a,k} - A_{X_a,k} w_{\zeta_*,k}\right|^2 \\
&= \frac{1}{c_{Tu}^2} \sum_{t,s=1}^{T} K_{h,tu} K_{h,su} e^{i(t-s)\lambda_k} \\
&\times \left[\mathbb{E}\left(X_{a,t,T} X_{a,s,T}\right) - \mathbb{E}\left(X_{a,t,T} \zeta_{*s}'\right) A_{X_a,k}^* - A_{X_a,k} \mathbb{E}\left(\zeta_{*t} X_{a,s,T}\right) + A_{X_a,k} \mathbb{E}\left(\zeta_{*t} \zeta_{*s}'\right) A_{X_a,k}^*\right].
\end{aligned}$$

Note that our bootstrap only replaces the innovation of error term. Suppose we decompose $\zeta_t = \left(\zeta_{X,t}', \zeta_{\varepsilon,t}\right)'$, then $\zeta_{*t} = \left(\zeta_{X,t}', \zeta_{*\varepsilon,t}\right)'$. Following the proof of Proposition A.1.3,

$\mathbb{E}\left(X_{a,t,T}X_{a,s,T}\right)$ stays the same as before, while

$$\mathbb{E}\left(X_{a,t,T}\zeta_{*s}^{'}\right)=\mathbb{E}\left(\sum_{k=0}^{\infty}A_{X_a,t,T}\left(k\right)\zeta_{t-k}\zeta_{*s}^{'}\right)=\sum_{k=0}^{\infty}\left(\widetilde{A}_{X_a,t,T}\left(k\right),0\right)\mathbb{E}\left(\begin{array}{cc}\zeta_{X,t-k}\zeta_{X,s}^{'} & \zeta_{X,t-k}\zeta_{*\varepsilon,s}\\ \zeta_{\varepsilon,t-k}\zeta_{X,s}^{'} & \zeta_{\varepsilon,t-k}\zeta_{*\varepsilon,s}\end{array}\right)$$

$$=\left(\widetilde{A}_{X_a,t,T}\left(t-s\right),0\right)\left(\begin{array}{cc}I_p & \mathbf{0}_{p\times 1}\\ \mathbf{0}_{1\times p} & 0\end{array}\right)=A_{X_a,t,T}\left(t-s\right)$$

by Assumption 1(i), which also replicates the result in the proof of Proposition A.1.3. And the same conclusion holds for $\mathbb{E}\left(\zeta_{*t}X_{a,s,T}\right)$. And $\mathbb{E}\left(\zeta_{*t}\zeta_{*s}^{'}\right)=\mathbb{E}\left(\zeta_{t}\zeta_{s}^{'}\right)$ by construction. Therefore $\mathbb{E}\left|w_{X_a,k}-A_{X_a,k}w_{\zeta_*,k}\right|^2$ is the same as $\mathbb{E}\left|w_{X_a,k}-A_{X_a,k}w_{\zeta,k}\right|^2$ as in the proof of Proposition A.1.4, and the same conclusion holds for $\mathbb{E}\left|w_{X_a,k}+A_{X_a,k}w_{\zeta_*,k}\right|^2$ as well. Next

$$\mathbb{E}\left|w_{\varepsilon_*,k}^{*}-w_{\zeta_*,k}^{*}A_{\varepsilon,k}^{*}\right|^2$$
$$=\frac{1}{c_{Tu}^2}\sum_{t,s=1}^{T}K_{h,tu}K_{h,su}e^{i(t-s)\lambda_k}\left[\mathbb{E}\left(\varepsilon_{*t,T}\varepsilon_{*s,T}\right)-\mathbb{E}\left(\varepsilon_{*t,T}\zeta_{*s}^{'}\right)A_{\varepsilon,k}^{*}-A_{\varepsilon,k}\mathbb{E}\left(\zeta_{*t}\varepsilon_{*s,T}\right)+A_{\varepsilon,k}\mathbb{E}\left(\zeta_{*t}\zeta_{*s}^{'}\right)A_{\varepsilon,k}^{*}\right],$$

where

$$\mathbb{E}\left(\varepsilon_{*t,T}\varepsilon_{*s,T}\right)=\mathbb{E}\left(\frac{2\pi}{T^2}\sum_{j=1}^{T}\sum_{k=1}^{T}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{s}{T},\lambda_k\right)}\sum_{r=1}^{T}\sum_{u=1}^{T}\zeta_{*\varepsilon,r}\zeta_{*\varepsilon,u}e^{i(r-t)\lambda_j}e^{-i(u-s)\lambda_k}\right)$$
$$=\frac{2\pi}{T^2}\sum_{j=1}^{T}\sum_{k=1}^{T}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{s}{T},\lambda_k\right)}\sum_{r=1}^{T}e^{ir\lambda_{j-k}}e^{-it\lambda_j}e^{is\lambda_k}$$
$$=\frac{2\pi}{T}\sum_{j=1}^{T}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{s}{T},\lambda_k\right)}e^{-i(t-s)\lambda_j}$$
$$=\frac{2\pi}{T}\sum_{j=1}^{T}\sqrt{f_{\varepsilon}\left(\frac{t}{T},\lambda_j\right)}\sqrt{f_{\varepsilon}\left(\frac{s}{T},\lambda_j\right)}e^{-i(t-s)\lambda_j}+o\left(1\right)$$
$$=\int_0^{2\pi}f_{\varepsilon}\left(u,\lambda\right)e^{-i(t-s)\lambda}d\lambda\cdot O\left(1+h\right)=\int_{-\pi}^{\pi}f_{\varepsilon}\left(u,\lambda\right)e^{-i(t-s)\lambda}d\lambda\cdot O\left(1+h\right)$$

where the third equality holds because $\sum_{r=1}^{T}e^{ir\lambda_{j-k}}$ is nonzero only when $j=k$, the fourth one holds by Assumption 8(ii), the fifth one holds by Riemann sum approximation and the corresponding smoothness conditions of $f_{\varepsilon}\left(u,\lambda\right)$ over $u\in(0,1)$, and the last equality holds by 2π-periodicity property of $f_{\varepsilon}\left(u,\lambda\right)e^{-i(t-s)\lambda}$. Note that again we obtain the same result as $\mathbb{E}\left(\varepsilon_{t,T}\varepsilon_{s,T}\right)$. Next

$$\mathbb{E}\left(\varepsilon_{*t,T}\zeta_{*s}^{'}\right)=\sqrt{\frac{2\pi}{T}}\sum_{j=1}^{T}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)}\frac{1}{\sqrt{T}}\sum_{r=1}^{T}\mathbb{E}\left(\zeta_{*\varepsilon,r}\zeta_{*s}^{'}\right)e^{i(r-t)\lambda_j},$$

where as before $\zeta_{*s}^{'}=\left(\zeta_{X,s}^{'},\zeta_{*\varepsilon,s}\right)^{'}$. Therefore it is sufficient to consider the element involving $\zeta_{*\varepsilon,s}$, which gives

$$\begin{aligned}
&\frac{\sqrt{2\pi}}{T}\sum_{j=1}^{T}\sqrt{\hat{f}_{\widehat{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)}\sum_{r=1}^{T}\mathbb{E}\left(\zeta_{*\varepsilon,r}\zeta_{*\varepsilon,s}\right)e^{i(r-t)\lambda_j}\\
&=\frac{\sqrt{2\pi}}{T}\sum_{j=1}^{T}\sqrt{f_{\varepsilon}\left(\frac{t}{T},\lambda_j\right)}e^{-i(t-s)\lambda_j}+o\left(1\right)\\
&=\frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}\sqrt{f_{\varepsilon}\left(u,\lambda\right)}e^{-i(t-s)\lambda}d\lambda\cdot O\left(1+h\right)\\
&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|A_{\varepsilon}\left(u,\lambda+\lambda_k\right)\right|e^{-i(t-s)\lambda+\lambda_k}d\lambda\cdot O\left(1+h\right),
\end{aligned}$$

which is also identical to $\mathbb{E}\left(\varepsilon_{t,T}\zeta_s^{'}\right)$ except imposing the norm to $\left|A_{\varepsilon}\left(u,\lambda+\lambda_k\right)\right|$, but it will have the same order in the end. And the same result holds for $\mathbb{E}\left(\zeta_{*t}\varepsilon_{*s,T}\right)$, which implies that $\mathbb{E}\left|w^*_{\varepsilon_*,k}-w^*_{\zeta_*,k}A^*_{\varepsilon,k}\right|^2$ has the same order as $\mathbb{E}\left|w^*_{\varepsilon,k}-w^*_{\zeta,k}A^*_{\varepsilon,k}\right|^2$, and the same conclusion holds for $\mathbb{E}\left|w^*_{\varepsilon_*,k}+w^*_{\zeta_*,k}A^*_{\varepsilon,k}\right|^2$. Together with what we have derived for $\mathbb{E}\left|w_{X_a,k}-A_{X_a,k}w_{\zeta_*,k}\right|^2$ and $\mathbb{E}\left|w_{X_a,k}+A_{X_a,k}w_{\zeta_*,k}\right|^2$, $\mathbb{E}\left|Q_1^*\right|$ has the same order as $\mathbb{E}\left|Q_1\right|$ in the proof of Proposition A.1.4.

Next for Q_2^*, we continue following the proof of Proposition A.1.4 and have that

$$\begin{aligned}
&\mathbb{E}\left|Q_2^*\right|^2\\
&=\mathbb{E}\left[\left(\sum_{k=l+1}^{j}\lambda_k^{d_{X_a}(u)+d_{\varepsilon}(u)}\left(I_{a\varepsilon_*,k}-A_{X_a,k}I_{\zeta_*,k}A^*_{\varepsilon,k}\right)\right)\left(\sum_{k=l+1}^{j}\lambda_k^{d_{X_a}(u)+d_{\varepsilon}(u)}\left(I^*_{a\varepsilon_*,k}-A_{\varepsilon,k}I_{\zeta_*,k}A^*_{X_a,k}\right)\right)\right]\\
&=\sum_{k=l+1}^{j}\lambda_k^{2(d_{X_a}(u)+d_{\varepsilon}(u))}\mathbb{E}\left[\left(w_{X_a,k}w^*_{\varepsilon_*,k}-A_{X_a,k}w_{\zeta_*,k}w^*_{\zeta_*,k}A^*_{\varepsilon,k}\right)\left(w_{\varepsilon_*,k}w^*_{X_a,k}-A_{\varepsilon,k}w_{\zeta_*,k}w^*_{\zeta_*,k}A^*_{X_a,k}\right)\right]\\
&+\sum_{s\neq k;k,s=l+1}^{j}\lambda_k^{d_{X_a}(u)+d_{\varepsilon}(u)}\lambda_s^{d_{X_a}(u)+d_{\varepsilon}(u)}\\
&\times\mathbb{E}\left[\left(w_{X_a,k}w^*_{\varepsilon_*,k}-A_{X_a,k}w_{\zeta_*,k}w^*_{\zeta_*,k}A^*_{\varepsilon,k}\right)\left(w_{\varepsilon_*,k}w^*_{X_a,k}-A_{\varepsilon,k}w_{\zeta_*,k}w^*_{\zeta_*,k}A^*_{X_a,k}\right)\right]\\
&\equiv Q^*_{21}+Q^*_{22},
\end{aligned}$$

where using the same notation as in the proof of Proposition A.1.4, we can denote

$$\begin{aligned}
Q^*_{21}=&\sum_{k=l+1}^{j}\lambda_k^{2(d_{X_a}(u)+d_{\varepsilon}(u))}\left[\mathbb{E}\left(a_ke_k^*\right)\mathbb{E}\left(e_ka_k^*\right)+\mathbb{E}\left(a_ke_k\right)\mathbb{E}\left(e_k^*a_k^*\right)+\mathbb{E}\left(a_ka_k^*\right)\mathbb{E}\left(e_k^*e_k\right)\right.\\
&-\mathbb{E}\left(A_kE_k^*\right)\mathbb{E}\left(e_ka_k^*\right)-\mathbb{E}\left(A_ke_k\right)\mathbb{E}\left(E_k^*a_k^*\right)-\mathbb{E}\left(A_ka_k^*\right)\mathbb{E}\left(E_k^*e_k\right)\\
&-\mathbb{E}\left(a_ke_k^*\right)\mathbb{E}\left(E_kA_k^*\right)-\mathbb{E}\left(a_kE_k\right)\mathbb{E}\left(e_k^*A_k^*\right)-\mathbb{E}\left(a_kA_k^*\right)\mathbb{E}\left(e_k^*E_k\right)\\
&\left.+\mathbb{E}\left(A_kE_k^*\right)\mathbb{E}\left(E_kA_k^*\right)+\mathbb{E}\left(A_kE_k\right)\mathbb{E}\left(E_k^*A_k^*\right)+\mathbb{E}\left(A_kA_k^*\right)\mathbb{E}\left(E_k^*E_k\right)\right]\\
&+\sum_{k=l+1}^{j}\lambda_k^{2(d_{X_a}(u)+d_{\varepsilon}(u))}\left[\text{cum}\left(a_k,e_k^*,e_k,a_k^*\right)-\text{cum}\left(A_k,E_k^*,e_k,a_k^*\right)\right.
\end{aligned}$$

$$-\mathrm{cum}\left(a_k, e_k^*, E_k, A_k^*\right) + \mathrm{cum}\left(A_k, E_k^*, E_k, A_k^*\right)]$$

$$\equiv Q_{21,a} + Q_{21,b},$$

where $Q_{21,a}$ is the part without cumulant and $Q_{21,b}$ is the part with cumulant, and

$$Q_{22} = \sum_{s\neq k;i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} [\mathbb{E}\left(a_k e_k^*\right)\mathbb{E}\left(e_s a_s^*\right) + \mathbb{E}\left(a_k e_s\right)\mathbb{E}\left(e_k^* a_s^*\right) + \mathbb{E}\left(a_k a_s^*\right)\mathbb{E}\left(e_k^* e_s\right)$$

$$- \mathbb{E}\left(A_k E_k^*\right)\mathbb{E}\left(e_s a_s^*\right) - \mathbb{E}\left(A_k e_s\right)\mathbb{E}\left(E_k^* a_s^*\right) - \mathbb{E}\left(A_k a_s^*\right)\mathbb{E}\left(E_k^* e_s\right)$$

$$- \mathbb{E}\left(a_k e_k^*\right)\mathbb{E}\left(E_s A_s^*\right) - \mathbb{E}\left(a_k E_s\right)\mathbb{E}\left(e_k^* A_s^*\right) - \mathbb{E}\left(a_k A_s^*\right)\mathbb{E}\left(e_k^* E_s\right)$$

$$+\mathbb{E}\left(A_k E_k^*\right)\mathbb{E}\left(E_s A_s^*\right) + \mathbb{E}\left(A_k E_s\right)\mathbb{E}\left(E_k^* A_s^*\right) + \mathbb{E}\left(A_k A_s^*\right)\mathbb{E}\left(E_k^* E_s\right)]$$

$$+ \sum_{s\neq k;i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)}$$

$$\times \left[\mathrm{cum}\left(a_k, e_k^*, e_s, a_s^*\right) - \mathrm{cum}\left(A_k, E_k^*, e_s, a_s^*\right) - \mathrm{cum}\left(a_k, e_k^*, E_s, A_s^*\right) + \mathrm{cum}\left(A_k, E_k^*, E_s, A_s^*\right)\right]$$

$$\equiv Q_{22,a} + Q_{22,b},$$

where $Q_{22,a}$ is the part without cumulant and $Q_{22,b}$ is the part with cumulant. Note that the order of $Q_{21,a}$ and $Q_{22,a}$, can be covered by analysis of $\mathbb{E}\left(e_k e_k^*\right)$, $\mathbb{E}\left(a_k a_k^*\right)$, $\mathbb{E}\left(A_k A_k^*\right)$ and $\mathbb{E}\left(E_k E_k^*\right)$, which can be dealt with identially as we did above, and the same order as in the proof of Proposition A.1.4 continues to hold. Thus it remains to consider the parts containing cumulants, $Q_{21,b}$ and $Q_{22,b}$, and it is sufficient to analyze $Q_{22,b}$ as it is dominant, which is given by

$$Q_{22,b} = \sum_{s\neq k;i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)}$$

$$\times \left[\mathrm{cum}\left(a_k, e_k^*, e_s, a_s^*\right) - \mathrm{cum}\left(A_k, E_k^*, e_s, a_s^*\right) - \mathrm{cum}\left(a_k, e_k^*, E_s, A_s^*\right) + \mathrm{cum}\left(A_k, E_k^*, E_s, A_s^*\right)\right]$$

$$= \sum_{s\neq k;i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)}$$

$$\times \left[\mathrm{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon}_*,k}, \widetilde{w}_{\widetilde{\varepsilon}_*,s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right) - \mathrm{cum}\left(A_{X_a,k}\widetilde{w}_{\widetilde{\zeta}_*,k}, \widetilde{w}^*_{\widetilde{\zeta}_*,k}A^*_{\varepsilon,k}, \widetilde{w}_{\widetilde{\varepsilon}_*,s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right)\right.$$

$$\left.-\mathrm{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon}_*,k}, A_{\varepsilon,s}\widetilde{w}_{\widetilde{\zeta}_*,s}, \widetilde{w}^*_{\widetilde{\zeta}_*,s}A_{X_a,s}\right) + \mathrm{cum}\left(A_{X_a,k}\widetilde{w}_{\widetilde{\zeta}_*,k}, \widetilde{w}^*_{\widetilde{\zeta}_*,k}A^*_{\varepsilon,k}, A_{\varepsilon,s}\widetilde{w}_{\widetilde{\zeta}_*,s}, \widetilde{w}^*_{\widetilde{\zeta}_*,s}A_{X_a,s}\right)\right],$$

where correspondingly we denote

$$\widetilde{w}_{\widetilde{X}_a,k} = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\widetilde{X}_{a,t,u,T}e^{it\lambda_k} = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\sqrt{\frac{2\pi T}{c_{Tu}^2}}X_{a,t,T}K_{h,tu}e^{it\lambda_k},$$

and denote $\widetilde{w}^*_{\widetilde{\varepsilon}_*,k}$ and $\widetilde{w}_{\widetilde{\zeta}_*,k}$ in the same manner. We consider $\mathrm{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon}_*,k}, \widetilde{w}_{\widetilde{\varepsilon}_*,s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right)$, which at large enough T is given by

$$\begin{aligned}
&\operatorname{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon}_*,k}, \widetilde{w}_{\widetilde{\varepsilon}_*,s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right) \\
&= \operatorname{cum}\left(\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\widetilde{X}_{a,t,u,T}e^{it\lambda_k}, \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\widetilde{\varepsilon}_{*t,u,T}e^{-it\lambda_k}, \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\widetilde{\varepsilon}_{*t,u,T}e^{it\lambda_s}, \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\widetilde{X}_{a,t,u,T}e^{-it\lambda_s}\right) \\
&= \frac{\sum_{t_1,\ldots,t_4=1}^{T}}{(2\pi T)^2}\operatorname{cum}\left(\widetilde{X}_{a,t_1,u,T}, \widetilde{\varepsilon}_{*t_2,u,T}, \widetilde{\varepsilon}_{*t_3,u,T}, \widetilde{X}_{a,t_4,u,T}\right)e^{i[(t_1-t_2)\lambda_k+(t_3-t_4)\lambda_s]} \\
&= \frac{\sum_{t_1,\ldots,t_4=1}^{T}}{(2\pi T)^2}\operatorname{cum}\left(\sum_{j=0}^{\infty}A_{X_a,t_1,T}(j)\,\zeta_{t_1-j,T}, \sum_{j=0}^{\infty}A_{\varepsilon,t_2,T}(j)\,\zeta_{t_2-j,T}, \sum_{j=0}^{\infty}A_{\varepsilon,t_3,T}(j)\,\zeta_{t_3-j,T}, \sum_{j=0}^{\infty}A_{X_a,t_4,T}(j)\,\zeta_{t_4-j,T}\right) \\
&\times\left(\frac{2\pi T}{c_{Tu}^2}\right)^2 K_{h,t_1u}K_{h,t_2u}K_{h,t_3u}K_{h,t_4u}e^{i[(t_1-t_2)\lambda_k+(t_3-t_4)\lambda_s]},
\end{aligned}$$

where the third equality holds by the fact that

$$\begin{aligned}
\varepsilon_{*t,T} &= \frac{\sqrt{2\pi}}{T}\sum_{j=1}^{T}\sqrt{\hat{f}_{\widetilde{\varepsilon}}\left(\frac{t}{T},\lambda_j\right)}\sum_{r=1}^{T}\zeta_{*\varepsilon,r}e^{i(r-t)\lambda_j} \\
&= \sum_{l=0}^{\infty}\int_{-\pi}^{\pi}|A_{\varepsilon}(u,\lambda)|\,e^{-il\lambda}d\lambda\zeta_{*\varepsilon,t-l} \equiv \sum_{l=0}^{\infty}A_*(u,l)\,\zeta_{*\varepsilon,t-l}
\end{aligned}$$

by denoting $r-t=-l$; and the second equality holds as we can replace $\zeta_{*\varepsilon,r}$ with its i.i.d. pairs to make the indices match, which requires a large T. We can see that $\varepsilon_{*t,T}$ can recapture the linear process structure asymptotically, and in (A.2.18) the part corresponding to $A_{\varepsilon}(u,\lambda)$ can be replaced by $|A_{\varepsilon}(u,\lambda)|$, which means the rest of the order is also identical to the one indicated in Proposition A.1.4, which then completes the proof of Theorem. ■

A.2 Proofs of the Technical Lemmas and Propositions

Proof of Lemma A.1.1. Denote the matrix of time-varying periodogram as $I_X(u,\lambda) \equiv w_X(u,\lambda)\,w_X^*(u,\lambda)$, and its (a,b)-th element as $I_{X_{ab}}(u,\lambda) = w_{X_a}(u,\lambda)\,w^*_{X_b}(u,\lambda)$ for any $1 \le a,b \le p$. We focus on the asymptotics of

$$\frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\operatorname{Re}\left(I_{X_{ab}}(u,\lambda_j)\right),$$

which is the (a,b)-th element of the first term in LHS of (A.1.1), and $1 \le a,b \le p$. To save space, we will suppress the dependence of $I_{X,ab}(u,\lambda_j)$, $I_{\zeta}(u,\lambda_j)$, $w_{X_a}(u,\lambda_j)$, $w_{\zeta}(u,\lambda_j)$, and $A_{X_a}(u,\lambda_j)$ on u and write them as $I_{X,ab,j}$, $I_{\zeta,j}$, $w_{X_a,j}$, $w_{\zeta,j}$, and $A_{X_a,j}$, respectively whenever no confusion can arise. Let

$$q_{1,j} = I_{X_{ab},j} - A_{X_a,j} I_{\zeta,j} A^*_{X_b,j} \text{ and } q_{2,j} = A_{X_a,j} I_{\zeta,j} A^*_{X_b,j} - f_{X_{ab}}(u, \lambda_j),$$

where $A_{X_a}(u, \lambda_j) \equiv \sum_{k=0}^{\infty} A^0_{X_a}(u, k) e^{ik\lambda_j}$ is defined as in (2.3.1). Consider the following decomposition:

$$\begin{aligned}
&\frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(I_{X_{ab}}(u, \lambda_j)\right) - \Gamma_{ab}(u) \\
&= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(q_{1,j}\right) + \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(q_{2,j}\right) \\
&+ \left\{ \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(f_{X_{ab}}(u, \lambda_j)\right) - \Gamma_{ab}(u) \right\} \\
&\equiv Q_{1,M} + Q_{2,M} + Q_{3,M}.
\end{aligned} \tag{A.2.1}$$

We prove the lemma by showing that $Q_{l,M} = o_p(1)$ for $l = 1, 2, 3$.

First, we study $Q_{1,M}$. By summation by parts, we have

$$\begin{aligned}
\frac{1}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(q_{1,j}\right) &= \sum_{j=1}^{M-1} \left(\lambda_j^{2\delta(u)} - \lambda_{j+1}^{2\delta(u)} \right) \frac{1}{M} \sum_{k=1}^{j} \mathrm{Re}\left(q_{1,k}\right) + \lambda_M^{2\delta(u)} \frac{1}{M} \sum_{j=1}^{M} \mathrm{Re}\left(q_{1,j}\right) \\
&\leq \sum_{j=1}^{M-1} \left| \lambda_j^{2\delta(u)} - \lambda_{j+1}^{2\delta(u)} \right| \frac{1}{M} \sum_{k=1}^{M} \left|\mathrm{Re}\left(q_{1,k}\right)\right| + \lambda_M^{2\delta(u)} \frac{1}{M} \sum_{j=1}^{M} \mathrm{Re}\left(q_{1,j}\right)
\end{aligned} \tag{A.2.2}$$

where

$$\begin{aligned}
&\sum_{j=1}^{M} \left|\mathrm{Re}\left(q_{1,j}\right)\right| \\
&\leq \sum_{j=1}^{M} \left| I_{X_{ab},j} - A_{X_a,j} I_{\zeta,j} A^*_{X_b,j} \right| = \sum_{j=1}^{M} \left| w_{X_a,j} w^*_{X_b,j} - A_{X_a,j} w_{\zeta,j} w^*_{\zeta,j} A^*_{X_b,j} \right| \\
&= \sum_{j=1}^{M} \frac{1}{2} \left| \left(w_{X_a,j} - A_{X_a,j} w_{\zeta,j}\right)\left(w^*_{X_b,j} + w^*_{\zeta,j} A^*_{X_b,j}\right) + \left(w_{X_a,j} + A_{X_a,j} w_{\zeta,j}\right)\left(w^*_{X_b,j} - w^*_{\zeta,j} A^*_{X_b,j}\right) \right| \\
&\leq \frac{1}{2} \sum_{j=1}^{M} \left| \left(w_{X_a,j} - A_{X_a,j} w_{\zeta,j}\right)\left(w^*_{X_b,j} + w^*_{\zeta,j} A^*_{X_b,j}\right) \right| + \frac{1}{2} \sum_{j=1}^{M} \left| \left(w_{X_a,j} + A_{X_a,j} w_{\zeta,j}\right)\left(w^*_{X_b,j} - w^*_{\zeta,j} A^*_{X_b,j}\right) \right| \\
&\equiv \frac{1}{2} \left(S_1(M) + S_2(M)\right).
\end{aligned}$$

By Cauchy-Schwarz inequality,

$$
\begin{aligned}
S_1(M) &\le \left(\sum_{j=1}^{M}\left|w_{X_a,j}-A_{X_a,j}w_{\zeta,j}\right|^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{M}\left|w^*_{X_b,j}+w^*_{\zeta,j}A^*_{X_b,j}\right|^2\right)^{\frac{1}{2}}\\
&\le \left(\sum_{j=1}^{M}\left|w_{X_a,j}-A_{X_a,j}w_{\zeta,j}\right|^2\right)^{\frac{1}{2}}\left(2\sum_{j=1}^{M}\left(I_{X_{bb},j}+A_{X_b,j}I_{\zeta,j}A^*_{X_b,j}\right)\right)^{\frac{1}{2}} \equiv S_{11}^{\frac{1}{2}}(M)\cdot S_{12}^{\frac{1}{2}}(M),
\end{aligned}
$$

and similarly

$$
S_2(M) \le \left(\sum_{j=1}^{M}\left|w_{X_b,j}-A_{X_b,j}w_{\zeta,j}\right|^2\right)^{\frac{1}{2}}\left(2\sum_{j=1}^{M}\left(I_{X_{aa},j}+A_{X_a,j}I_{\zeta,j}A^*_{X_a,j}\right)\right)^{\frac{1}{2}} \equiv S_{21}^{\frac{1}{2}}(M)\cdot S_{22}^{\frac{1}{2}}(M).
$$

By Proposition A.1.3, we have

$$
S_{11}(M) = o_p\left(TF_{X_{aa}}(u,\lambda_M)\right) \text{ and } S_{12}(M) = o_p\left(TF_{X_{aa}}(u,\lambda_M)\right).
$$

It follows that $S_1(M) = o_p\left(TF^{\frac{1}{2}}_{X_{aa}}(u,\lambda_m)F^{\frac{1}{2}}_{X_{bb}}(u,\lambda_m)\right)$. Similarly,

$$
S_2(M) = o_p\left(TF^{\frac{1}{2}}_{X_{aa}}(u,\lambda_M)F^{\frac{1}{2}}_{X_{bb}}(u,\lambda_M)\right).
$$

Then $\sum_{j=1}^{M}\mathrm{Re}(q_{1,j}) = o_p\left(TF^{\frac{1}{2}}_{X_{aa}}(u,\lambda_M)F^{\frac{1}{2}}_{X_{bb}}(u,\lambda_M)\right)$, and the same order holds for $\left|\sum_{k=1}^{M} Re\left(q_{1,k}\right)\right|$. This, in conjunction with (A.2.2), implies that

$$
\begin{aligned}
\frac{1}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(q_{1,j}\right) \le \frac{1}{M}\sum_{j=1}^{M-1}\left|\lambda_j^{2\delta(u)}-\lambda_{j+1}^{2\delta(u)}\right| o_p\left(TF^{\frac{1}{2}}_{X_{aa}}(u,\lambda_M)F^{\frac{1}{2}}_{X_{bb}}(u,\lambda_M)\right)\\
+\frac{1}{M}\lambda_M^{2\delta(u)}o_p\left(TF^{\frac{1}{2}}_{X_{aa}}(u,\lambda_M)F^{\frac{1}{2}}_{X_{bb}}(u,\lambda_M)\right) \equiv B_{1,M}+B_{2,M}.
\end{aligned}
\tag{A.2.3}
$$

Noting that $\lambda_{j+1} = \frac{2\pi(j+1)}{T} = \frac{2\pi j}{T}(1+\frac{1}{j}) = \lambda_j(1+\frac{1}{j})$, we have for some absolute constant C that

$$
\left|\lambda_j^{2\delta(u)}-\lambda_{j+1}^{2\delta(u)}\right| = \lambda_j^{2\delta(u)}\left|\left(1+\frac{1}{j}\right)^{2\delta(u)}-1\right| \le \frac{C}{j}\lambda_j^{2\delta(u)} = O\left(\frac{1}{2\pi j}\lambda_j^{2\delta(u)}\right) = O\left(\frac{1}{T}\lambda_j^{2\delta(u)-1}\right),
$$

and note that the existence of C can be justified by the leading terms of binomial series $\left(1+\frac{1}{j}\right)^{2\delta(u)}$ and the fact that the remaining terms are of smaller magnitude as j goes from 1 to M.

Then by Assumption 2, $B_{1,M}$ follows that

$$\begin{aligned}
B_{1,M} &= \frac{1}{M}\sum_{j=1}^{M-1}\left|\lambda_j^{2\delta(u)} - \lambda_{j+1}^{2\delta(u)}\right| o_p\left(TF_{X_{aa}}^{\frac{1}{2}}(u,\lambda_M) F_{X_{bb}}^{\frac{1}{2}}(u,\lambda_M)\right) \\
&= \frac{1}{M} O\left(\frac{1}{T}\sum_{j=1}^{M-1}\lambda_j^{2\delta(u)-1}\right) o_p\left(T\lambda_M^{1-d_{X_a}(u)-d_{X_b}(u)}\right) \\
&= \frac{T}{M} o_p\left(\lambda_M^{2\delta(u)+1-d_{X_a}(u)-d_{X_b}(u)}\right) = o_p\left(\lambda_M^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\right).
\end{aligned}$$

Similarly

$$B_{2,M} = \frac{1}{M}\lambda_M^{2\delta(u)} o_p\left(TF_{X_{aa}}^{\frac{1}{2}}(u,\lambda_M) F_{X_{bb}}^{\frac{1}{2}}(u,\lambda_M)\right) = o_p\left(\lambda_M^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\right).$$

It follows that $\frac{1}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}(q_{1,j}) = o_p\left(\lambda_M^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\right)$ and therefore by Riemann sum approximation,

$$Q_{1,M} = \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(q_{1,j}\right) = o_p(1).$$

Next, we study $Q_{2,M}$. We decompose it as

$$\begin{aligned}
Q_{2,M} &= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X_a,j}I_{\zeta,j}A_{X_b,j}^* - f_{X_{ab}}(u,\lambda_j)\right) \\
&= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X_a,j}\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t^{'}K_{h,tu}^2 A_{X_b,j}^* - f_{X_{ab}}(u,\lambda_j)\right) \\
&+ \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X_a,j}\frac{1}{c_{Tu}^2}\sum_{t\neq s}\zeta_t\zeta_s^{'}K_{h,tu}K_{h,su}e^{i(t-s)\lambda_j}A_{X_b,j}^*\right) \\
&\equiv \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(q_{2,j1}+q_{2,j2}\right) \\
&\equiv Q_{2,M1} + Q_{2,M2}.
\end{aligned}$$

In the following we study $Q_{2,M1}$ and $Q_{2,M2}$ separately. Firstly for $Q_{2,M1}$ we have its expectation given by

$$\begin{aligned}
\mathbb{E}\left(Q_{2,M1}\right) &= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X_a,j}\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\mathbb{E}\left(\zeta_t\zeta_t'\right)K_{h,tu}^2 A_{X_b,j}^* - f_{X_{ab}}(u,\lambda_j)\right) \\
&= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}K_{h,tu}^2 A_{X_a,j}A_{X_b,j}^* - f_{X_{ab}}(u,\lambda_j)\right) \\
&= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(\frac{1}{2\pi}A_{X_a,j}A_{X_b,j}^* - f_{X_{ab}}(u,\lambda_j)\right) = 0
\end{aligned}$$

by Assumption 1. Then for the variance of $Q_{2,M1}$ we have

$$\begin{aligned}\operatorname{Var}\left(Q_{2,M1}\right) &= \frac{\lambda_M^{2[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)]}}{M^2}\sum_{j=1}^{M}\sum_{k=1}^{M}\lambda_j^{2\delta(u)}\lambda_k^{2\delta(u)}\operatorname{Cov}\left(\operatorname{Re}\left(q_{2,j1}\right),\operatorname{Re}\left(q_{2,k1}\right)\right)\\ &\leq \frac{\lambda_M^{2[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)]}}{M^2}\sum_{j=1}^{M}\sum_{k=1}^{M}\lambda_j^{2\delta(u)}\lambda_k^{2\delta(u)}\left\{\operatorname{Var}\left(\operatorname{Re}\left(q_{2,j1}\right)\right)\operatorname{Var}\left(\operatorname{Re}\left(q_{2,k1}\right)\right)\right\}^{1/2}.\end{aligned}$$

Note that

$$\begin{aligned}\operatorname{Var}\left(\operatorname{Re}\left(q_{2,j1}\right)\right) &= \mathbb{E}\left[\operatorname{Re}\left(A_{X_a,j}\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2A_{X_b,j}^* - f_{X_{ab}}\left(u,\lambda_j\right)\right)\right]^2\\ &\leq \mathbb{E}\left|A_{X_a,j}\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2A_{X_b,j}^* - f_{X_{ab}}\left(u,\lambda_j\right)\right|^2\\ &= \mathbb{E}\left|A_{X_a,j}\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2A_{X_b,j}^* - \frac{1}{2\pi}A_{X_a,j}A_{X_b,j}^*\right|^2 \equiv \mathbb{E}\left|A_{X_a,j}D_{1j}A_{X_b,j}^*\right|^2\end{aligned}$$

where we denote $D_{1j} \equiv D_1\left(u,\lambda_j\right) \equiv \frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2 - \frac{1}{2\pi}I_{p+1}$, which is a real symmetric matrix. Then following the above equalities, and using the fact that $\operatorname{tr}(ABCD) = \operatorname{vec}\left(A'\right)'\times\left(D'\otimes B\right)\times\operatorname{vec}\left(C\right)$, we have

$$\begin{aligned}\mathbb{E}\left|A_{X_a,j}D_{1j}A_{X_b,j}^*\right|^2 &= \mathbb{E}\left(A_{X_a,j}D_{1j}A_{X_b,j}^*A_{X_b,j}D_{1j}A_{X_a,j}^*\right) = \mathbb{E}\left[\operatorname{tr}\left(A_{X_a,j}^*A_{X_a,j}D_{1j}A_{X_b,j}^*A_{X_b,j}D_{1j}\right)\right]\\ &= \operatorname{vec}\left(A_{X_a,j}^*A_{X_a,j}\right)'\mathbb{E}\left(D_{1j}\otimes D_{1j}\right)\operatorname{vec}\left(A_{X_b,j}^*A_{X_b,j}\right),\end{aligned}$$

Let $D_{1j,mn}$ denote the (m,n)-th element of D_{1j}:

$$D_{1j,mn} = \frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_{t,m}\zeta_{t,n}K_{h,tu}^2 - \frac{\mathbf{1}\left(m=n\right)}{2\pi},$$

we have that

$$\begin{aligned}\mathbb{E}\left(D_{1j}\otimes D_{1j}\right) &= \mathbb{E}\left[\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2 - \frac{1}{2\pi}I_{p+1}\right)\otimes\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2 - \frac{1}{2\pi}I_{p+1}\right)\right]\\ &= \mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\right) - \frac{1}{2\pi}I_{p+1}\otimes\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\right)\\ &- \mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\right)\otimes\frac{1}{2\pi}I_{p+1} + \frac{1}{\left(2\pi\right)^2}\left(I_{p+1}\otimes I_{p+1}\right)\\ &= \mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\right) - \frac{1}{\left(2\pi\right)^2}\left(I_{p+1}\otimes I_{p+1}\right),\end{aligned}$$

where the last equality holds by the fact that

$$\frac{1}{2\pi}I_{p+1}\otimes\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\right)=\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\right)\otimes\frac{1}{2\pi}I_{p+1}$$

$$=\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}K_{h,tu}^2 I_{p+1}\otimes\frac{1}{2\pi}I_{p+1}=\frac{1}{(2\pi)^2}\left(I_{p+1}\otimes I_{p+1}\right).$$

Next we can see that

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\otimes\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\zeta_t\zeta_t'K_{h,tu}^2\right) \qquad \text{(A.2.4)}\\
&=\frac{1}{c_{Tu}^4}\sum_{t,s=1}^{T}\mathbb{E}\left(\zeta_t\zeta_t'\otimes\zeta_s\zeta_s'\right)K_{h,tu}^2K_{h,su}^2\\
&\le\frac{1}{c_{Tu}^4}\sum_{t,s=1}^{T}\left[C\left(I_{p+1}\otimes I_{p+1}\right)\mathbf{1}\left(t=s\right)+\left(I_{p+1}\otimes I_{p+1}\right)\mathbf{1}\left(t\neq s\right)\right]K_{h,tu}^2K_{h,su}^2\\
&=\frac{1}{c_{Tu}^4}\left[C\left(I_{p+1}\otimes I_{p+1}\right)\sum_{t=1}^{T}K_{h,tu}^4+\left(I_{p+1}\otimes I_{p+1}\right)\sum_{t\neq s}^{T}K_{h,tu}^2K_{h,su}^2\right]\\
&\le\frac{\left(I_{p+1}\otimes I_{p+1}\right)C\sum_{t=1}^{T}K_{h,tu}^4}{c_{Tu}^4}+\frac{1}{(2\pi)^2}I_{p+1}\otimes I_{p+1}\\
&=O\left(\frac{1}{Th}\right)I_{p+1}\otimes I_{p+1}+\frac{1}{(2\pi)^2}I_{p+1}\otimes I_{p+1}
\end{aligned}$$

where the inequality holds by Assumption 1(i) and the last equality holds by Assumption 3(ii) and Riemann sum approximation that $\sum_{t=1}^{T}K_h^4\left(t-Tu\right)=Th^{-3}\kappa_{04}\left[1+o\left(1\right)\right]$ and $\sum_{t=1}^{T}K_h^2\left(t-Tu\right)=Th^{-1}\kappa_{02}\left[1+o\left(1\right)\right]$. Then in sum, $\mathbb{E}\left(D_{1j}\otimes D_{1j}\right)=O\left(\frac{1}{Th}\right)I_{p+1}\otimes I_{p+1}$, which then implies that

$$\begin{aligned}
Var\left(\mathrm{Re}\left(q_{2,j1}\right)\right)&=\mathbb{E}\left|A_{X_a,j}D_{1j}A_{X_b,j}^*\right|^2=O\left(\frac{1}{Th}\right)\left|A_{X_a,j}A_{X_b,j}^*\right|^2\\
&=O\left(\frac{1}{Th}\lambda_j^{-2d_{X_a}(u)-2d_{X_b}(u)}\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left(Q_{2,M1}\right)&\le\frac{\lambda_M^{2[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)]}}{M^2}\sum_{j=1}^{M}\sum_{k=1}^{M}\lambda_j^{2\delta(u)}\lambda_k^{2\delta(u)}\left\{\mathrm{Var}\left(\mathrm{Re}\left(q_{2,j1}\right)\right)\mathrm{Var}\left(\mathrm{Re}\left(q_{2,k1}\right)\right)\right\}^{1/2}\\
&=\frac{\lambda_M^{2[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)]}}{M^2Th}\sum_{j=1}^{M}\sum_{k=1}^{M}O\left(\lambda_j^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\lambda_k^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\right)\\
&=O\left(\frac{1}{Th}\right)=o\left(1\right)
\end{aligned}$$

by Assumption 4(ii) and 4(iii), and thus $Q_{2,M1}=o_p\left(1\right)$.

Next for $Q_{2,M2}$, we can rewrite it as

$$\begin{aligned}
Q_{2,M2} &= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X_a,j}\frac{1}{c_{Tu}^2}\sum_{t\neq s}\zeta_t\zeta_s' K_{h,tu}K_{h,su}e^{i(t-s)\lambda_j}A_{X_b,j}^*\right)\\
&= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\frac{1}{c_{Tu}^2}\sum_{t\neq s}\zeta_t'\mathrm{Re}\left(\sum_{j=1}^{M}\lambda_j^{2\delta(u)}A_{X_a,j}'\overline{A}_{X_b,j}K_{h,tu}K_{h,su}e^{i(t-s)\lambda_j}\right)\zeta_s\\
&\equiv \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}}{M}\frac{1}{c_{Tu}^2}\mathrm{Re}\left(\sum_{t\neq s}\zeta_t'\Phi_{t,s,M}\zeta_s\right).
\end{aligned}$$

Then it is easy to see that $\mathbb{E}(Q_{2,M2})=0$, and its order can be determined by $\mathbb{E}\left|Q_{2,M2}\right|^2$, which is further bounded by

$$\begin{aligned}
\mathbb{E}\left|Q_{2,M2}\right|^2 &\leq \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\mathbb{E}\left|\sum_{t\neq s}\zeta_t'\Phi_{t,s,M}\zeta_s\right|^2\\
&= \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t_1\neq s_1}\sum_{t_2\neq s_2}\mathbb{E}\left(\zeta_{t_1}'\Phi_{t_1,s_1,M}\zeta_{s_1}\zeta_{t_2}'\overline{\Phi}_{t_2,s_2,M}\zeta_{s_2}\right)\\
&= \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t_1\neq s_1}\sum_{t_2\neq s_2}\mathbb{E}\left[\mathrm{tr}\left(\zeta_{s_1}'\Phi_{t_1,s_1,M}'\zeta_{t_1}\zeta_{t_2}'\overline{\Phi}_{t_2,s_2,M}\zeta_{s_2}\right)\right]\\
&= \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t_1\neq s_1}\sum_{t_2\neq s_2}\mathbb{E}\left[\mathrm{tr}\left(\Phi_{t_1,s_1,M}'\zeta_{t_1}\zeta_{t_2}'\overline{\Phi}_{t_2,s_2,M}\zeta_{s_2}\zeta_{s_1}'\right)\right]\\
&= \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t_1\neq s_1}\sum_{t_2\neq s_2}\mathrm{vec}\left(\overline{\Phi}_{t_2,s_2,M}\right)'\mathbb{E}\left(\zeta_{s_2}\zeta_{s_1}'\otimes\zeta_{t_2}\zeta_{t_1}'\right)\mathrm{vec}\left(\Phi_{t_1,s_1,M}\right).
\end{aligned} \tag{A.2.5}$$

And by Assumption 1(i) and (2.3.6),

$$\mathbb{E}\left(\zeta_{s_2}\zeta_{s_1}'\otimes\zeta_{t_2}\zeta_{t_1}'\right)=\left(I_{p+1}\otimes I_{p+1}\right)\mathbf{1}\left(t_1=t_2\neq s_2=s_1\right)+\widetilde{P}\mathbf{1}\left(t_1=s_2\neq t_2=s_1\right)$$

with sparse permutation matrix $\widetilde{P}$ defined in Assumption 1(i) and (2.3.6). Then (A.2.5) follows that

$$\begin{aligned}
&\frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t_1\neq s_1}\sum_{t_2\neq s_2}\mathrm{vec}\left(\overline{\Phi}_{t_2,s_2,M}\right)'\mathbb{E}\left(\zeta_{s_2}\zeta_{s_1}'\otimes\zeta_{t_2}\zeta_{t_1}'\right)\mathrm{vec}\left(\Phi_{t_1,s_1,M}\right)\\
&= \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t\neq s}\mathrm{vec}\left(\overline{\Phi}_{t,s,M}\right)'\left(I_{p+1}\otimes I_{p+1}\right)\mathrm{vec}\left(\Phi_{t,s,M}\right)\\
&+ \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t\neq s}\mathrm{vec}\left(\overline{\Phi}_{s,t,M}\right)'\widetilde{P}\mathrm{vec}\left(\Phi_{t,s,M}\right)\\
&\equiv V_1+V_2,
\end{aligned}$$

where firstly V_1 is further given by

$$
\begin{aligned}
|V_1| &= \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2 c_{Tu}^4} \left| \sum_{t\neq s} \text{vec}\left(\overline{\Phi}_{t,s,M}\right)' \left(I_{p+1}\otimes I_{p+1}\right) \text{vec}\left(\Phi_{t,s,M}\right) \right| \\
&= \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2 c_{Tu}^4} \left| \sum_{t\neq s} \text{vec}\left(\sum_{j=1}^{M} \lambda_j^{2\delta(u)} A_{X_a,j}^{*} A_{X_b,j} K_{h,tu} K_{h,su} e^{-i(t-s)\lambda_j} \right)' \right. \\
&\left. \times \left(I_{p+1}\otimes I_{p+1}\right) \text{vec}\left(\sum_{j=1}^{M} \lambda_j^{2\delta(u)} A'_{X_a,j} \overline{A}_{X_b,j} K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j} \right) \right| \\
&\leq \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2} \sum_{j,k=1}^{M} \lambda_j^{2\delta(u)} \lambda_k^{2\delta(u)} \left| \frac{1}{c_{Tu}^4} \sum_{t\neq s} K_{h,tu}^2 K_{h,su}^2 e^{-i(t-s)\lambda_{j-k}} \right| \\
&\times \left\| \text{vec}\left(A_{X_a,j}^{*} A_{X_b,j}\right) \right\| \left\| I_{p+1}\otimes I_{p+1} \right\| \left\| \text{vec}\left(A'_{X_a,k} \overline{A}_{X_b,k}\right) \right\| \\
&\leq \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2} \sum_{j=1}^{M} \lambda_j^{4\delta(u)} \left| \frac{1}{c_{Tu}^4} \sum_{t\neq s} K_{h,tu}^2 K_{h,su}^2 \right| \\
&\times \left\| \text{vec}\left(A_{X_a,j}^{*} A_{X_b,j}\right) \right\| \left\| I_{p+1}\otimes I_{p+1} \right\| \left\| \text{vec}\left(A'_{X_a,j} \overline{A}_{X_b,j}\right) \right\| \\
&+ \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2} \sum_{j\neq k} \lambda_j^{2\delta(u)} \lambda_k^{2\delta(u)} \left| \frac{1}{c_{Tu}^4} \sum_{t\neq s} K_{h,tu}^2 K_{h,su}^2 e^{-i(t-s)\lambda_{j-k}} \right| \\
&\times \left\| \text{vec}\left(A_{X_a,j}^{*} A_{X_b,j}\right) \right\| \left\| I_{p+1}\otimes I_{p+1} \right\| \left\| \text{vec}\left(A'_{X_a,k} \overline{A}_{X_b,k}\right) \right\| \\
&\equiv V_{11} + V_{12},
\end{aligned}
$$

where the norm $\|V\|$ for a K-dimensional complex vector V is defined by $\sqrt{\sum_{i=1}^{K} |V_i|^2}$, with $|\cdot|^2$ being the squared-modulus of a complex number. Then for V_{11} we have

$$
V_{11} = \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{2\pi M^2} \sum_{j=1}^{M} \lambda_j^{4\delta(u)} \left\| \text{vec}\left(A_{X_a,j}^{*} A_{X_b,j}\right) \right\| \left\| I_{p+1}\otimes I_{p+1} \right\| \left\| \text{vec}\left(A'_{X_a,j} \overline{A}_{X_b,j}\right) \right\|,
$$

using the fact that

$$
\frac{1}{c_{Tu}^4} \sum_{t\neq s} K_{h,tu}^2 K_{h,su}^2 \leq \frac{1}{c_{Tu}^4} \sum_{t,s=1}^{T} K_{h,tu}^2 K_{h,su}^2 = \frac{1}{2\pi}.
$$

And V_{12} is given by

$$
\begin{aligned}
V_{12} &= O\left(h^{-1}\right) \frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2} \\
&\times \sum_{j\neq k; |j-k|\leq \delta(u)} \lambda_j^{2\delta(u)} \lambda_k^{2\delta(u)} \left\| \text{vec}\left(A_{X_a,j}^{*} A_{X_b,j}\right) \right\| \left\| I_{p+1}\otimes I_{p+1} \right\| \left\| \text{vec}\left(A'_{X_a,k} \overline{A}_{X_b,k}\right) \right\|
\end{aligned}
$$

using the fact that

$$\left|\frac{1}{c_{Tu}^4}\sum_{t\neq s}K_{h,tu}^2K_{h,su}^2e^{-i(t-s)\lambda_{j-k}}\right| \leq \left|\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}K_{h,tu}^2e^{it\lambda_{j-k}}\right|^2$$

$$= \left|\frac{\sum_{t=1}^{T}k^2\left(\frac{t-Tu}{Th}\right)e^{it\lambda_{j-k}}}{2\pi Th\kappa_{02}+o(1)}\right|^2 = O\left(h^{-1}\right)$$

accroding to Assumption 3(iv). Suppose we denote the k-th element of $A_{X_a,j}$ as $A_{X_a,j(k)}$, then the norm $\left\|\text{vec}\left(A_{X_a,j}^*A_{X_a,j}\right)\right\|$ is further given by

$$\left\|\text{vec}\left(A_{X_a,j}^*A_{X_b,j}\right)\right\| = \left[\sum_{k,l=1}^{p+1}\left|\overline{A}_{X_a,j(k)}A_{X_b,j(l)}\right|^2\right]^{\frac{1}{2}} = \left[\sum_{k,l=1}^{p+1}\left|A_{X_a,j(k)}\right|^2\left|A_{X_b,j(l)}\right|^2\right]^{\frac{1}{2}}$$

$$= \left(\left|A_{X_a,j}\right|^2\left|A_{X_b,j}\right|^2\right)^{\frac{1}{2}} = \left|A_{X_a,j}\right|\left|A_{X_b,j}\right| = O\left(\lambda_j^{-d_{X_a}(u)-d_{X_b}(u)}\right),$$

which implies that

$$V_{11} = \frac{\lambda_M^{2[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)]}}{2\pi M^2}\sum_{j=1}^{M}O\left(\lambda_j^{4\delta(u)-2d_{X_a}(u)-2d_{X_b}(u)}\right) = O\left(\frac{1}{M}\right) = o(1),$$

and

$$V_{12} = O\left(\frac{\lambda_M^{2\left[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)\right]}}{M^2h}\sum_{j\neq k;|j-k|\leq\delta(u)}\lambda_j^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\lambda_k^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\right)$$

$$= O\left(\frac{1}{Mh}\right) = o(1)$$

by Assumption 4(iii). This concludes the proof of asymptotic negligibility of V_1.

Next for V_2 we have in norm that

$$|V_2| = \left|\frac{\lambda_M^{2[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)]}}{M^2}\frac{1}{c_{Tu}^4}\sum_{t\neq s}\text{vec}\left(\overline{\Phi}_{s,t,M}\right)'\widetilde{P}\text{vec}\left(\Phi_{t,s,M}\right)\right|$$

$$\leq \frac{\lambda_M^{2[d_{X_a}(u)+d_{X_b}(u)-2\delta(u)]}}{M^2}\sum_{j,k=1}^{M}\lambda_j^{2\delta(u)}\lambda_k^{2\delta(u)}\left|\frac{1}{c_{Tu}^4}\sum_{t\neq s}K_{h,tu}^2K_{h,su}^2e^{-i(t-s)\lambda_{j-k}}\right|$$

$$\times\left\|\text{vec}\left(A_{X_a,j}^*A_{X_b,j}\right)\right\|\left\|\widetilde{P}\right\|\left\|\text{vec}\left(A_{X_a,k}'\overline{A}_{X_b,k}\right)\right\|.$$

Note that the formula above is identical to the one for V_1 except $\|I_{p+1}\otimes I_{p+1}\|$ is replaced by $\left\|\widetilde{P}\right\|$. And by definition of $\widetilde{P}$, as presented in Assumption 1(i) and (2.3.6), the Frobenius norm of permutation matrix $\left\|\widetilde{P}\right\|$ is identical to an identity matrix of the same dimension. Therefore the order of V_2 shall be the same as that of V_1. Therefore combining what we have so far, we prove that $Q_{2,M2} = o_p(1)$.

Now, we study $Q_{3,M}$. By Assumption 2 and the numerical property of Riemann sum approximation, we have

$$Q_{3,M} = \lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)} \frac{1}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(f_{X,ab}\left(u,\lambda_j\right)\right) - \frac{G_{X,ab}}{1-d_{X_a}\left(u\right)-d_{X_b}\left(u\right)+2\delta\left(u\right)}$$
$$\asymp G_{X,ab}\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)} \frac{1}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)} - \frac{G_{X,ab}}{1-d_{X_a}\left(u\right)-d_{X_b}\left(u\right)+2\delta\left(u\right)}$$
$$= o\left(1\right)$$

where the last equality follows because

$$\lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)} \frac{1}{M}\sum_{j=1}^{M}\lambda_j^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}$$
$$= \lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}\left(\frac{2\pi M}{T}\right)^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\frac{1}{M}\sum_{j=1}^{M}\left(\frac{j}{M}\right)^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}$$
$$= \lambda_M^{d_{X_a}(u)+d_{X_b}(u)-2\delta(u)}\left(\frac{2\pi M}{T}\right)^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}\left[\int_0^1 x^{2\delta(u)-d_{X_a}(u)-d_{X_b}(u)}dx + O\left(M^{-1}\right)\right]$$
$$= \frac{1}{1-d_{X_a}\left(u\right)-d_{X_b}\left(u\right)+2\delta\left(u\right)} + O\left(M^{-1}\right).$$

In sum, we have shown that $Q_{1,M}+Q_{2,M}+Q_{3,M} = o_p\left(1\right)$. This completes the proof of Lemma A.1.1. ■

Proof of Lemma A.1.2. Let $\eta = \left(\eta_1,...,\eta_p\right)'$ be an arbitrary nonrandom $p\times 1$ real vector such that $\|\eta\| = 1$. By Cramér-Wold Device, it suffices to consider

$$\sum_{a=1}^{p}\eta_a\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{\frac{h}{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(I_{a\varepsilon}\left(u,\lambda_j\right)\right) \equiv B_M,$$

where $I_{a\varepsilon}\left(u,\lambda_j\right) = w_{X,a}\left(u,\lambda_j\right)w_\varepsilon^*\left(u,\lambda_j\right)$ denotes the cross-periodogram. As before, we will suppress the dependence of $I_{a\varepsilon}\left(u,\lambda_j\right)$, $w_{X_a}\left(u,\lambda_j\right)$, $w_\zeta\left(u,\lambda_j\right)$, $A_{X_a}\left(u,\lambda_j\right)$, $A_\varepsilon\left(u,\lambda_j\right)$ and $I_\zeta\left(u,\lambda_j\right)$ on u and write them as $I_{a\varepsilon,j}$, $w_{X_a,j}$, $w_{\zeta,j}$, $A_{X_a,j}$, $A_{\varepsilon,j}$ and $I_{\zeta,j}$. To proceed, we consider the following decomposition

$$B_M = \sum_{a=1}^{p}\eta_a\frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(I_{a\varepsilon,j}-A_{X_a,j}I_{\zeta,j}A_{\varepsilon,j}^*\right)$$
$$+\sum_{a=1}^{p}\eta_a\frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A_{X_a,j}I_{\zeta,j}A_{\varepsilon,j}^*\right)$$
$$\equiv B_{1,M}+B_{2,M}. \tag{A.2.6}$$

We shall prove the asymptotic negligibility of $B_{1,M}$ and an establish a CLT for $B_{2,M}$.

First, we study $B_{1,M}$ that is given by

$$B_{1,M} = \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(I_{a\varepsilon,j} - A_{X_a,j} I_{\zeta,j} A_{\varepsilon,j}^*\right)$$
$$\equiv \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}} \widetilde{B}_{1,Ma},$$

then by summation by parts, for any $a = 1, \ldots, p$,

$$\widetilde{B}_{1,Ma} = \sum_{j=1}^{M-1} \left(\lambda_j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)} - \lambda_{j+1}^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\right) \sum_{k=1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \mathrm{Re}\left(I_{a\varepsilon,k} - A_{X_a,k} I_{\zeta,k} A_{\varepsilon,k}^*\right)$$
$$+ \lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)} \sum_{j=1}^{M} \lambda_j^{d_{X_a}(u)+d_\varepsilon(u)} \mathrm{Re}\left(I_{a\varepsilon,j} - A_{X_a,j} I_{\zeta,j} A_{\varepsilon,j}^*\right)$$
$$= O_p\left(\lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)} \left(h^{-1} (\log M)^2 + h^{-1} M^{\frac{1}{4}} (\log M)^{\frac{1}{2}} + h^{-1} M^{\frac{1}{2}} T^{-\frac{1}{4}} (\log M)^{\frac{1}{2}}\right)\right)$$

by Proposition A.1.4, which implies that

$$B_{1,M} = \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}} \widetilde{B}_{1,Ma}$$
$$= \sum_{a=1}^{p} \eta_a O_p\left(\frac{(\log M)^2}{M^{\frac{1}{2}} h^{\frac{1}{2}}} + \frac{(\log M)^{\frac{1}{2}}}{M^{\frac{1}{4}} h^{\frac{1}{2}}} + \frac{(\log M)^{\frac{1}{2}}}{T^{\frac{1}{4}} h^{\frac{1}{2}}}\right) = o_p(1)$$

by Assumption 4(iii).

Next, we study $B_{2,M}$. We make the following decomposition

$$B_{2,M} = \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left(A_{X_a,j} I_{\zeta,j} A_{\varepsilon,j}^*\right)$$
$$= \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left[A_{X_a,j} \frac{\sum_{t=1}^{T} \zeta_t \zeta_t' K_{h,tu}^2}{c_{Tu}^2} A_{\varepsilon,j}^*\right]$$
$$+ \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left[A_{X_a} \frac{\sum_{t\neq s} \zeta_t \zeta_s' K_{h,tu} K_{h,su} e^{i(t-s)\lambda_j}}{c_{Tu}^2} A_\varepsilon^*\right]$$
$$\equiv B_{2,M1} + B_{2,M2}.$$

For $B_{2,M1}$, we make further decomposition:

$$B_{2,M1} = \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left[A_{X_a,j} \frac{\sum_{t=1}^{T} \left(\zeta_t \zeta_t' - I_{p+1}\right) K_{h,tu}^2}{c_{Tu}^2} A_{\varepsilon,j}^*\right]$$
$$+ \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}}{M} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \mathrm{Re}\left[\frac{1}{2\pi} A_{X_a,j} A_{\varepsilon,j}^*\right]$$
$$= o_p(1)$$

by how we study $Q_{2,M1}$ in the proof of Lemma A.1.1 and Assumption 2(iv).

To study $B_{2,M2}$, note that

$$B_{2,M2} = \sum_{t=1}^{T} \frac{\zeta_t' K_{h,tu}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} K_{h,tu}^2}} \sum_{s<t} C_{t-s,T}(u) \frac{\zeta_s K_{h,su}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} K_{h,tu}^2}} \equiv \sum_{t=1}^{T} Z_{t,T}(u) \tag{A.2.7}$$

where $C_{t,T}(u)$ is given by

$$C_{t,T}(u) = \sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{2\pi T\sqrt{M}} \sum_{j=1}^{M} \lambda_j^{2\delta(u)} \text{Re}\left(A_{X_a,j}'\overline{A}_{\varepsilon,j} + A_{\varepsilon,j}'\overline{A}_{X_a,j}\right)\cos(t\lambda_j) \tag{A.2.8}$$

$$\equiv \frac{\sqrt{h}}{\sqrt{M}2\pi T}\sum_{j=1}^{M}\theta_j \cos(t\lambda_j)$$

Apparently, $Z_{t,T}(u)$ is a martingale difference sequence under Assumption 1. Then by the martigale central limit theorem (see, e.g., Hall and Heyde (1980, Ch3.2), Robinson (1995a), Lobato (1999), Nielsen (2005)), it is sufficient to prove:

(i) $\sum_{t=1}^{T} \mathbb{E}\left(Z_{t,T}^2(u) \mid \mathcal{F}_{t-1}\right) \xrightarrow{p} \sum_{a=1}^{p}\sum_{b=1}^{p} \eta_a\eta_b\Omega_{ab}$ for some $p \times p$ matrix Ω;

(ii) $\sum_{t=1}^{T} \mathbb{E}\left(Z_{t,T}^2(u)\mathbf{1}(|Z_{t,T}(u)| > \varepsilon)\right) \xrightarrow{p} 0$ for all $\varepsilon > 0$.

By Proposition A.1.5, these two conditions hold. Consequently, we have

$$\sum_{a=1}^{p} \eta_a \frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{\sqrt{M}} \sum_{j=1}^{M} \text{Re}\left(A_{X_a,j} I_{\zeta,j} A_{\varepsilon,j}^*\right) \xrightarrow{d} \mathcal{N}\left(0, \sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\Omega_{ab}(u)\right)$$

where $\Omega_{ab}(u) = \Theta^*\left(\frac{G_{X,aa}^{\frac{1}{2}} G_{X,bb}^{\frac{1}{2}} G_\varepsilon}{1-d_{X_a}(u)-d_{X_b}(u)-2d_\varepsilon(u)+4\delta(u)}\right)$. And together with $B_{1,M} = o_p(1)$ shown just now, we complete the proof of Lemma A.1.2. ■

Proof of Proposition A.1.3. (i) The proof follows closely from that of Proposition 4 in Lobato (1997). The major difference lies on the presence of kernel weight in the definitions of w_{X_a} and w_ζ. By the nonnegativity of $S_a(M)$, its stochastic order can be obtained by the order of its expectation by Markov's inequality. Recall that $w_{X_a,j} = w_{X_a}(u,\lambda_j)$, $w_{\zeta,j} = w_\zeta(u,\lambda_j)$, and $A_{X_a,j} = A_{X_a,j}(u,\lambda_j)$. We have

$$\begin{aligned}
\mathbb{E}[S_a(M)] &= \mathbb{E}\left(\frac{1}{T}\sum_{j=1}^{M}\left(w_{X_a,j} - A_{X_a,j}w_{\zeta,j}\right)\left(w_{X_a,j}^* - w_{\zeta,j}^* A_{X_a,j}^*\right)\right) \\
&= \mathbb{E}\left[\frac{1}{Tc_{Tu}^2}\sum_{j=1}^{M}\sum_{t=1}^{T}\left[X_{a,t,T} - A_{X_a,j}\zeta_t\right] K_{h,tu} e^{it\lambda_j} \sum_{s=1}^{T}\left[X_{a,s,T}' - \zeta_s' A_{X_a,j}^*\right] K_{h,su} e^{-is\lambda_j}\right] \\
&= \frac{1}{Tc_{Tu}^2}\sum_{j=1}^{M}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{h,tu}K_{h,su}e^{i(t-s)\lambda_j}
\end{aligned}$$

$$\cdot\left[\mathbb{E}\left(X_{a,t,T}X'_{a,s,T}\right)-\mathbb{E}\left(X_{a,t,T}\zeta'_{s}\right)A^{*}_{X_a,j}-A_{X_a,j}\mathbb{E}\left(\zeta_t X_{a,s,T}\right)+A_{X_a,j}\mathbb{E}\left(\zeta_t\zeta'_s\right)A^{*}_{X_a,j}\right].$$

We will evaluate $\mathbb{E}\left(X_{a,t,T}X'_{a,s,T}\right)$, $\mathbb{E}\left(X_{a,t,T}\zeta'_s\right)$, $\mathbb{E}\left(\zeta_t X'_{a,s,T}\right)$, and $\mathbb{E}\left(\zeta_t\zeta'_s\right)$. Let $\psi_1 = T^{-1} + \left|u-\frac{t}{T}\right| + \left|u-\frac{s}{T}\right|$ and $\psi_{2,t} = T^{-1} + \left|u-\frac{t}{T}\right|$. For $\mathbb{E}\left(X_{a,t,T}X'_{a,s,T}\right)$, we have by Assumption 1(ii) that

$$\begin{aligned}
\mathbb{E}\left(X_{a,t,T}X'_{a,s,T}\right) &= \mathbb{E}\left(\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}A_{X_a,t,T}(k)\,\zeta_{t-k}\zeta'_{s-l}A'_{X_a,s,T}(l)\right) = \sum_{l=0}^{\infty}A_{X_a,t,T}(l+(t-s))\,A'_{X_a,s,T}(l)\\
&= \int_{-\pi}^{\pi} f_{X_a}(u,\lambda)\,e^{-i(t-s)\lambda}d\lambda\cdot O(1+\psi_1)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} A_{X_a}(u,\lambda)\,A^{*}_{X_a}(u,\lambda)\,e^{-i(t-s)\lambda}d\lambda\cdot O(1+\psi_1)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} A_{X_a}(u,\lambda+\lambda_j)\,A^{*}_{X_a}(u,\lambda+\lambda_j)\,e^{-i(t-s)(\lambda+\lambda_j)}d\lambda\cdot O(1+\psi_1)
\end{aligned}$$

where the third equality holds by the approximation in (2.3.7), the fourth equality holds by the definition of the spectral density, and the fifth equality holds because both $A_{X_a}(u,\lambda)$ and $e^{-i(t-s)\lambda}$ have period 2π along the argument λ. Similarly,

$$\begin{aligned}
\mathbb{E}\left(X_{a,t,T}\zeta'_s\right) &= \mathbb{E}\left(\sum_{k=0}^{\infty}A_{X_a,t,T}(k)\,\zeta_{t-k}\zeta'_s\right) = A_{X_a,t,T}(t-s) = A_{X_a}(u,t-s)\cdot O(1+\psi_{2,t})\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} A_{X_a}(u,\lambda+\lambda_j)\,e^{-i(t-s)(\lambda+\lambda_j)}d\lambda\cdot O(1+\psi_{2,t})
\end{aligned}$$

where the second equality holds by Assumption 1(i), and the third equality holds by the approximations in (2.3.1) and (2.3.2) in Assumption 1(ii). By the same token,

$$\begin{aligned}
\mathbb{E}\left(\zeta_t X'_{a,s,T}\right) &= \mathbb{E}\left(\zeta_t\sum_{k=0}^{\infty}\zeta'_{s-k}A'_{X_a,s,T}(k)\right) = A'_{X_a,s,T}(s-t) = A'_{X_a}(u,s-t)\cdot O(1+\psi_{2,s})\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} A^{*}_{X_a}(u,\lambda+\lambda_j)\,e^{-i(t-s)(\lambda+\lambda_j)}d\lambda\cdot O(1+\psi_{2,s}).
\end{aligned}$$

In addition, it is easy to see that $\mathbb{E}\left(\zeta_t\zeta_s^{'}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-i(t-s)(\lambda+\lambda_j)}d\lambda$. Then

$$\begin{aligned}
&\mathbb{E}\left[S_a(M)\right]\\
&= \frac{1}{Tc^2_{Tu}}\sum_{j=1}^{M}\sum_{t=1}^{T}\sum_{s=1}^{T}K_{h,tu}K_{h,su}e^{i(t-s)\lambda_j}\\
&\times\left[\int_{-\pi}^{\pi}\left[A_{X_a}(u,\lambda+\lambda_j)-A_{X_a}(u,\lambda_j)\right]\left[A^{*}_{X_a}(u,\lambda+\lambda_j)-A^{*}_{X_a}(u,\lambda_j)\right]e^{-i(t-s)(\lambda+\lambda_j)}d\lambda\cdot O(1+\psi_1)\right]\\
&= \frac{1}{(2\pi)^2T}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\widetilde{K}(u,\lambda)\left[A_{X_a}(u,\lambda+\lambda_j)-A_{X_a}(u,\lambda_j)\right]\left[A^{*}_{X_a}(u,\lambda+\lambda_j)-A^{*}_{X_a}(u,\lambda_j)\right]d\lambda\cdot O(1)\\
&+\frac{1}{Tc^2_{Tu}}\sum_{j=1}^{M}\sum_{t=1}^{T}\sum_{s=1}^{T}K_{h,tu}K_{h,su}e^{i(t-s)\lambda_j}\int_{-\pi}^{\pi}\left[A_{X_a}(u,\lambda+\lambda_j)-A_{X_a}(u,\lambda_j)\right]
\end{aligned}$$

$$\times \left[A^*_{X_a}(u,\lambda+\lambda_j) - A^*_{X_a}(u,\lambda_j)\right] e^{-i(t-s)(\lambda+\lambda_j)} d\lambda \cdot O(\psi_1)$$
$$\equiv ES_1 + ES_2, \tag{A.2.9}$$

where $\widetilde{K}(u,\lambda)$ is a Fejér kernel "weighted " by $K_h(t-Tu)/\sqrt{\sum_{t=1}^T K_h^2(t-Tu)}$, as

$$\widetilde{K}(u,\lambda) = \frac{1}{\sum_{t=1}^T K_h^2(t-Tu)} \left|\sum_{t=1}^T K_h(t-Tu) e^{it\lambda}\right|^2 \tag{A.2.10}$$

It is easy to argue that $ES_2 = o(ES_1)$ by using the fact that $\left|u - \frac{t}{T}\right| \leq h$ on the support of kernel $K_h(t-Tu)$. So we focus on the derivation of the order of ES_1 below.

To derive the order of ES_1, we first derive that the order of $\widetilde{K}(u,\lambda)$. Note that by summation by parts and triangle inequality,

$$\begin{aligned}
\left|\sum_{t=1}^T K_h(t-Tu) e^{it\lambda}\right| &= \left|\sum_{t=1}^{T-1} [K_h(t-Tu) - K_h(t+1-Tu)] \sum_{s=1}^t e^{is\lambda} + K_h(T-Tu) \sum_{t=1}^T e^{it\lambda}\right| \\
&\leq \left|\sum_{t=1}^{T-1} (K_{h,tu} - K_{h,t+1,u}) \sum_{s=1}^t e^{is\lambda}\right| + K_{h,Tu} \left|\sum_{t=1}^T e^{it\lambda}\right| \\
&\leq \sum_{t=1}^{T-1} |K_{h,tu} - K_{h,t+1,u}| \max_{1\leq t\leq T} \left|\sum_{s=1}^t e^{is\lambda}\right| + K_{h,Tu} \left|\sum_{t=1}^T e^{it\lambda}\right|.
\end{aligned}$$

As in Hansen (2008, pp. 740-741), Assumption 3 implies that

$$|K_{h,tu} - K_{h,t+1,u}| = h^{-1} \left|k\left(\frac{t-Tu}{Th}\right) - k\left(\frac{t+1-Tu}{Th}\right)\right| \leq h^{-1} \frac{1}{Th} k^*\left(\frac{t-Tu}{Th}\right),$$

where $k^*(u) = \Lambda_1 \mathbf{1}(|u| \leq 2)$. Then

$$\sum_{t=1}^{T-1} |K_{h,tu} - K_{h,t+1,u}| \leq h^{-1} \frac{1}{Th} \sum_{t=1}^{T-1} k^*\left(\frac{t-Tu}{Th}\right) \leq Ch^{-1} \text{ for some } C < \infty.$$

Next, as Wang (2019) observes, $\sin(\lambda/2) \geq \lambda/\pi$ in the interval $(0,\pi)$ because the function $g(\lambda) = \sin(\lambda/2) - \lambda/\pi$ is concave on $(0,\pi)$ and $f(0) = f(\pi) = 0$. Similarly, $\sin(\lambda/2) \leq \lambda/\pi$ for $\lambda \in (-\pi, 0)$. It follows that for all $\lambda \in (-\pi,\pi)$, we have $|\sin(\lambda/2)| \geq |\lambda|/\pi$, and

$$\begin{aligned}
\left|\sum_{s=1}^t e^{is\lambda}\right| &= \left|\frac{e^{i\lambda}\left(1-e^{it\lambda}\right)}{1-e^{i\lambda}}\right| = \left|e^{i\lambda} e^{-i\lambda/2} e^{it\lambda/2} \frac{e^{-it\lambda/2} - e^{it\lambda/2}}{e^{-i\lambda/2} - e^{i\lambda/2}}\right| \\
&= \left|e^{i(t+1)\lambda/2} \frac{-2i\sin(t\lambda/2)}{-2i\sin(\lambda/2)}\right| \leq \frac{C}{|\lambda|} \text{ uniformly in } t.
\end{aligned}$$

In addition, $K_{h,Tu} = h^{-1}k\left(\frac{T-Tu}{Th}\right) \leq Ch^{-1}$. Consequently, we have

$$\left|\sum_{t=1}^T K_h(t-Tu) e^{it\lambda}\right| \leq \frac{C}{h|\lambda|}$$

where recall the constant C can vary across lines. This result, in conjunction with the fact that $\sum_{t=1}^{T} K_h^2(t-Tu) = Th^{-1}\int k(u)^2 du[1+o(1)]$, implies that

$$\widetilde{K}(u,\lambda) = \frac{1}{\sum_{t=1}^{T} K_h^2(t-Tu)} \left|\sum_{t=1}^{T} K_h(t-Tu) e^{it\lambda}\right|^2 = \frac{1}{Th^{-1}} O\left(h^{-2}|\lambda|^{-2}\right) = O\left(T^{-1}h^{-1}|\lambda|^{-2}\right). \tag{A.2.11}$$

Given the above order of $\widetilde{K}(u,\lambda)$, we are ready to derive the order of ES_1. By definition of spectral density in Assumption 2, we have for some constants C_1 and C_2,

$$\begin{aligned} ES_1 &\leq \frac{1}{2\pi T}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\widetilde{K}(u,\lambda)\left[C_1 f_{X_{aa}}(u,\lambda+\lambda_j) + C_2 f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\ &\leq C\frac{1}{2\pi T}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j) + f_{X_{aa}}(u,\lambda_j)\right]d\lambda \equiv C\cdot\overline{ES}_1. \end{aligned}$$

Following the proof of Proposition 4 in Robinson (1994b), we make the following decomposition for $\overline{ES}_1$:

$$\begin{aligned} \overline{ES}_1 &= \frac{1}{2\pi T}\sum_{j=1}^{M}\left[\left(\int_{-\pi}^{-\tau} + \int_{\tau}^{\pi}\right) + \left(\int_{-\tau}^{-h^{-1}\lambda_m} + \int_{h^{-1}\lambda_m}^{\tau}\right) + \left(\int_{-h^{-1}\lambda_m}^{-\lambda_m} + \int_{\lambda_m}^{h^{-1}\lambda_m}\right)\right. \\ &\quad \left. + \left(\int_{-\lambda_m}^{-\varepsilon_T\lambda_m} + \int_{\varepsilon_T\lambda_m}^{\lambda_m}\right) + \int_{-\varepsilon_T\lambda_m}^{\varepsilon_T\lambda_m}\right]\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j) + f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\ &\equiv ES_{11} + ES_{12} + ES_{13} + ES_{14} + ES_{15} \end{aligned}$$

where τ is some fixed but small enough constant and ε_T is a shrinking sequence to be defined later. We then study ES_{1l}'s, $l = 1, \ldots, 5$ in turn.

First, for ES_{11}, we have

$$\begin{aligned} ES_{11} &= \frac{1}{T}\sum_{j=1}^{M}\left(\int_{-\pi}^{-\tau} + \int_{\tau}^{\pi}\right)\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j) + f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\ &\leq \frac{C}{T^2h}\sum_{j=1}^{M}\left(\int_{-\pi}^{-\tau} + \int_{\tau}^{\pi}\right)\lambda^{-2} f_{X_{aa}}(u,\lambda+\lambda_j)\,d\lambda + \frac{C}{T^2h}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j)\int_{\tau}^{\pi}\lambda^{-2}d\lambda \\ &\leq \frac{C}{T^2h\tau^2}\sum_{j=1}^{M}\int_{-\pi}^{\pi} f_{X_{aa}}(u,\lambda+\lambda_j)\,d\lambda + \frac{C}{T^2h\tau}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j) \\ &= \frac{C}{T^2h\tau^2}\sum_{j=1}^{M}\int_{-\pi}^{\pi} f_{X_{aa}}(u,\lambda)\,d\lambda + \frac{C}{T^2h\tau}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j) \\ &= O\left(\frac{M}{T^2h} + \frac{1}{Th}F_{X_{aa}}(u,\lambda_M)\right) = O\left(\frac{1}{Th}F_{X_{aa}}(u,\lambda_M)\right) \end{aligned}$$

where the first inequality holds by the order of $\widetilde{K}(u,\lambda)$ derived above and the fact that $\widetilde{K}(u,\lambda)$ is an even function, the second inequality holds by the non-negativity of spectral density function $f_{X_{aa}}(u,\lambda)$, and the second equality holds by the property that spectral density function $f_{X_{aa}}(u,\lambda)$ has period 2π, the next to last equality follows from the fact that $\frac{1}{T}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j) = O\left(F_{X_{aa}}(u,\lambda_M)\right)$ by Proposition 1 in Robinson (1994b), and the last equality holds by the fact that $M = o(T)$. Note that the above tricks will be applied repeatedly in the following proof of this proposition.

Next, we study ES_{12}. Note that

$$
\begin{aligned}
ES_{12} &= \frac{1}{T}\sum_{j=1}^{M}\left(\int_{-\tau}^{-h^{-1}\lambda_M} + \int_{h^{-1}\lambda_M}^{\tau}\right)\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j) + f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\
&\leq \frac{C}{T^2 h}\sum_{j=1}^{M}\frac{1}{(h^{-1}\lambda_M)^2}\left(\int_{-\tau}^{-h^{-1}\lambda_M} + \int_{h^{-1}\lambda_M}^{\tau}\right) f_{X_{aa}}(u,\lambda+\lambda_j)\,d\lambda \\
&+ \frac{C}{T^2 h}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j)\int_{h^{-1}\lambda_M}^{\tau}\frac{1}{\lambda^2}d\lambda \\
&\leq \frac{Ch}{T^2\lambda_M^2}\sum_{j=1}^{M}\int_{h^{-1}\lambda_M}^{\tau} f_{X_{aa}}(u,\lambda_j)\,d\lambda + \frac{C}{T^2\lambda_M}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j) \\
&= O\left(\frac{1}{T^2\lambda_M}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j)\right) = O\left(\frac{1}{T\lambda_M}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j)\right) = O\left(\frac{1}{M}F_{X_{aa}}(u,\lambda_M)\right)
\end{aligned}
$$

where the first inequality holds by the even property of $\widetilde{K}(u,\lambda)$ and its bound derived above, and the second inequality holds because $f_{X_{aa}}(u,\lambda+\lambda_j) \leq f_{X_{aa}}(u,\lambda_j)$ for all $|\lambda| \in \left[h^{-1}\lambda_M,\tau\right)$ for small fixed τ by the singularity of spectral density around zero frequency.

Next, for ES_{13} we have

$$
\begin{aligned}
ES_{13} &= \frac{1}{T}\sum_{j=1}^{M}\left(\int_{-h^{-1}\lambda_M}^{-\lambda_M} + \int_{\lambda_M}^{h^{-1}\lambda_M}\right)\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j) + f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\
&\leq \frac{C}{T^2\lambda_M^2 h}\sum_{j=1}^{M}\int_{\lambda_M}^{h^{-1}\lambda_M} f_{X_{aa}}(u,\lambda+\lambda_j)\,d\lambda + \frac{C}{T^2 h}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j)\int_{\lambda_M}^{h^{-1}\lambda_M}\frac{1}{\lambda^2}d\lambda \\
&\leq \frac{C}{T^2\lambda_M^2 h}\sum_{j=1}^{M}\int_{\lambda_M}^{h^{-1}\lambda_M} f_{X_{aa}}(u,\lambda_j)\,d\lambda + \frac{C}{T^2 h}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j)\int_{\lambda_M}^{h^{-1}\lambda_M}\frac{1}{\lambda^2}d\lambda \\
&= O\left(\frac{h^{-1}\lambda_M}{T^2\lambda_M^2 h}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j) + \frac{1}{T^2\lambda_M h}\sum_{j=1}^{M} f_{X_{aa}}(u,\lambda_j)\right) \\
&= O\left(\left(\frac{1}{T\lambda_M h^2} + \frac{1}{T\lambda_M h}\right)F_{X_{aa}}(u,\lambda_M)\right) = O\left(\frac{1}{Mh^2}F_{X_{aa}}(u,\lambda_M)\right).
\end{aligned}
$$

To study ES_{14}, we specify a sequence ε_T such that as $T\to\infty$, $\varepsilon_T \to 0$ and $\varepsilon_T M \to \infty$, and there exists a constant C_ε such that $C_\varepsilon \varepsilon_T M$ is an integer. Under this specification, we have

$$
\begin{aligned}
ES_{14} &= \frac{1}{T}\sum_{j=1}^{M}\left(\int_{-\lambda_M}^{-\varepsilon_T\lambda_M}+\int_{\varepsilon_T\lambda_M}^{\lambda_M}\right)\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j)+f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\
&\leq \frac{C}{T}\sum_{j=1}^{M}\int_{\varepsilon_T\lambda_M}^{\lambda_M}\widetilde{K}(\lambda)\,f_{X_{aa}}(u,\lambda+\lambda_j)\,d\lambda+\frac{C}{T}\sum_{j=1}^{M}f_{X_{aa}}(u,\lambda_j)\int_{\varepsilon_T\lambda_M}^{\lambda_M}\widetilde{K}(\lambda)\,d\lambda \\
&\leq \frac{C}{T^2h}\sum_{j=1}^{M}f_{X_{aa}}(u,\lambda_j)\int_{\varepsilon_T\lambda_M}^{\lambda_M}\frac{1}{\lambda^2}d\lambda \\
&= O\left(\frac{1}{M\varepsilon_T h}F_{X_{aa}}(u,\lambda_M)\right). \qquad \text{(A.2.12)}
\end{aligned}
$$

Finally, for ES_{15} we make the following decomposition:

$$
\begin{aligned}
ES_{15} &= \frac{1}{T}\sum_{j=1}^{C_\varepsilon\varepsilon_T M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j)+f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\
&+\frac{1}{T}\sum_{j=C_\varepsilon\varepsilon_T M+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j)+f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\
&\equiv ES_{15,1}+ES_{15,2}.
\end{aligned}
$$

For $ES_{15,1}$, we have

$$
\begin{aligned}
ES_{15,1} &= \frac{1}{T}\sum_{j=1}^{C_\varepsilon\varepsilon_T M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j)+f_{X_{aa}}(u,\lambda_j)\right]d\lambda \\
&\leq \frac{2}{T}\sum_{j=1}^{C_\varepsilon\varepsilon_T M}f_{X_{aa}}(u,\lambda_j)\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\,d\lambda \leq \frac{4\pi}{T}\sum_{j=1}^{C_\varepsilon\varepsilon_T M}f_{X_{aa}}(u,\lambda_j) \\
&= O\left(F_{X_{aa}}(u,C_\varepsilon\varepsilon_T\lambda_M)\right)=O\left(\varepsilon_T^{1-2d_{X_a}(u)}F_{X_{aa}}(u,\lambda_M)\right),
\end{aligned}
$$

where the first inequality holds by the nonnegativity of $\widetilde{K}(u,\lambda)$ and montone property of $f_{X_{aa}}(u,\cdot)$ in the neighborhood of 0 for each u, and the second inequality holds by the fact that

$$
\begin{aligned}
\int_{-\pi}^{\pi}\widetilde{K}(u,\lambda)\,d\lambda &= \frac{1}{\sum_{t=1}^{T}K_h^2(t-Tu)}\int_{-\pi}^{\pi}\left|\sum_{t=1}^{T}K_h(t-Tu)\,e^{it\lambda}\right|^2 d\lambda \\
&= \frac{1}{\sum_{t=1}^{T}K_h^2(t-Tu)}\sum_{t=1}^{T}\sum_{s=1}^{T}K_h(t-Tu)\,K_h(s-Tu)\int_{-\pi}^{\pi}e^{i(t-s)\lambda}d\lambda \\
&= \frac{\sum_{t=1}^{T}K_h^2(t-Tu)}{\sum_{t=1}^{T}K_h^2(t-Tu)}\int_{-\pi}^{\pi}d\lambda = 2\pi.
\end{aligned}
$$

Here we use the fact that $\int_{-\pi}^{\pi}e^{i(t-s)\lambda}d\lambda=0$ for all $t\neq s$ with $t,s=1,...,T$. The last equality holds by the fact that in the neighborhood of zero frequency, $F_{X_{aa}}(u,\lambda)=O\left(\lambda^{1-2d_{X_a}(u)}\right)$

for each u by Assumption 2(i).

For $ES_{15,2}$ we can consider instead the following term

$$\begin{aligned}
\overline{ES}_{15,2} &= \frac{1}{T}\sum_{j=C_\varepsilon\varepsilon_T M+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left|A_{X_a}(u,\lambda+\lambda_j)-A_{X_a}(u,\lambda_j)\right|^2 d\lambda \\
&\leq \frac{C}{T}\sum_{j=C_\varepsilon\varepsilon_T M+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left|(\lambda+\lambda_j)^{-d_{X_a}(u)}-\lambda_j^{-d_{X_a}(u)}\right|^2 d\lambda \\
&= O\left(\frac{1}{T}\sum_{j=C_\varepsilon\varepsilon_T M+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left|\lambda_j^{-d_{X_a}(u)-1}\lambda\right|^2 d\lambda\right) \\
&= O\left(\frac{1}{T}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\lambda^2\widetilde{K}(u,\lambda)\,d\lambda\sum_{j=C_\varepsilon\varepsilon_T M+1}^{M}\lambda_j^{-2d_{X_a}(u)-2}\right) \\
&= O\left(\frac{1}{Th}\lambda_{\varepsilon_T M}^{-2d_{X_a}(u)-1}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}d\lambda\right) = O\left(\frac{\varepsilon_T\lambda_M}{Th}(\varepsilon_T\lambda_M)^{-2d_{X_a}(u)-1}\right) \\
&= O\left(\frac{1}{Mh\varepsilon_T^{2d_{X_a}(u)}}F_{X_{aa}}(u,\lambda_M)\right) \qquad \text{(A.2.13)}
\end{aligned}$$

where the first equality follows from the fact that $|A_{X_a}(u,\lambda_j)| = O(f_{X_{aa}}^{\frac{1}{2}}(u,\lambda_j))$, and the second equality holds by first-order Taylor expansion of $(\lambda+\lambda_j)^{-d_{X_a}(u)}$ at $\lambda=0$. Then we can express $ES_{15,2} = \frac{1}{2\pi}\overline{ES}_{15,2} + \left(ES_{15,2} - \frac{1}{2\pi}\overline{ES}_{15,2}\right)$, and note that

$$\begin{aligned}
\frac{1}{2\pi}\overline{ES}_{15,2} &= \frac{1}{T}\sum_{j=C_\varepsilon\varepsilon_T m+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\frac{1}{2\pi}\left|A_{X_a}(u,\lambda+\lambda_j)-A_{X_a}(u,\lambda_j)\right|^2 d\lambda \\
&= \frac{1}{T}\sum_{j=C_\varepsilon\varepsilon_T m+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_j)+f_{X_{aa}}(u,\lambda_j)\right] \\
&\quad - \frac{1}{2\pi T}\sum_{j=C_\varepsilon\varepsilon_T m+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left[A_{X_a}(u,\lambda+\lambda_j)A_{X_a}^*(u,\lambda_j)+A_{X_a}(u,\lambda_j)A_{X_a}^*(u,\lambda+\lambda_j)\right] \\
&\equiv ES_{15,2} + \frac{1}{2\pi}\overline{RES}_{15,2},
\end{aligned}$$

and $\overline{RES}_{15,2}$ is further given by

$$\begin{aligned}
\left|\overline{RES}_{15,2}\right| &\leq \frac{1}{2\pi T}\sum_{j=C_\varepsilon\varepsilon_T m+1}^{M}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(u,\lambda)\left|A_{X_a}(u,\lambda+\lambda_j)A_{X_a}^*(u,\lambda_j)+A_{X_a}(u,\lambda_j)A_{X_a}^*(u,\lambda+\lambda_j)\right| \\
&= O\left(\frac{1}{2\pi T}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\widetilde{K}(\lambda)\sum_{j=C_\varepsilon\varepsilon_T m+1}^{M}\lambda_j^{-2d_{X_a}(u)}\right) \\
&\leq O\left(\lambda_M^{1-2d_{X_a}(u)}\frac{1}{Th}\int_{-\varepsilon_T\lambda_M}^{\varepsilon_T\lambda_M}\frac{1}{\lambda^2}d\lambda\right) = O\left(\frac{1}{Mh\varepsilon_T}F_{X_{aa}}(u,\lambda_M)\right).
\end{aligned}$$

Then we can conclude that $ES_{15,2} = O\left(\frac{1}{Mh\varepsilon_T}F_{X_{aa}}(u,\lambda_M)\right)$.

Note that one appropriate choice of ε_T can be $\varepsilon_T = h^\delta$ where $\delta = \frac{1}{1-2\overline{d}_X}$ is defined as in Assumption 4(iii), which then makes both (A.2.12) and (A.2.13) $O\left(hF_{X_{aa}}\left(u,\lambda_M\right)\right)$, and thus

$$\begin{aligned}\mathbb{E}\left(S_a\left(M\right)\right) &= O\left(\left(\frac{1}{Mh^2}+\varepsilon_T^{1-2d_{X_a}(u)}+\frac{1}{Mh\varepsilon_T}\right)F_{X_{aa}}\left(u,\lambda_M\right)\right)\\ &= O\left(hF_{X_{aa}}\left(u,\lambda_M\right)\right)\end{aligned}$$

by choices of ε_T above and by Assumption 4(iii). This is similar to the conclusion of Proposition 4 in Robinson (1994b). And thus we finish the proof of argument (i).

(ii) By the same reasoning as used in the proof of (i), we can prove $\mathbb{E}\left[\frac{1}{T}\sum_{j=1}^{M}\left(I_{X,aa}+A_{X_a}I_{\zeta}A_{X_a}^*\right)\right]=O\left(hF_{X_{aa}}\left(u,\lambda_M\right)\right)$. The result follows by the Markov inequality. ■

Proof of Proposition A.1.4. Using the notation in the proof of Lemma A.1.1, we have

$$\begin{aligned}\widetilde{B}_{1,Ma} &= \sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left[I_{a\varepsilon,j}-A_{X_a,j}I_{\zeta,j}A_{\varepsilon,j}^*\right] \qquad \text{(A.2.14)}\\ &= \sum_{j=1}^{M-1}\left(\lambda_j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}-\lambda_{j+1}^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\right)\sum_{k=1}^{j}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathrm{Re}\left(I_{a\varepsilon,k}-A_{X_a,k}I_{\zeta,k}A_{\varepsilon,k}^*\right)\\ &+ \lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\sum_{j=1}^{M}\lambda_j^{d_{X_a}(u)+d_\varepsilon(u)}\mathrm{Re}\left(I_{a\varepsilon,j}-A_{X_a,j}I_{\zeta,j}A_{\varepsilon,j}^*\right)\\ &\equiv \sum_{j=1}^{M-1}\left(\lambda_j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}-\lambda_{j+1}^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\right)\cdot\widetilde{Q}\left(j\right)+\lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\cdot\widetilde{Q}\left(M\right).\end{aligned}$$

In the following we try to apply the proof of (C.2) in Lobato (1999) to derive the order for both $\widetilde{Q}\left(j\right)$ for arbitrary fixed $1\leq j<M$ and $\widetilde{Q}\left(M\right)$. Take $\widetilde{Q}\left(j\right)$ for example, it has the decomposition given by

$$\begin{aligned}\widetilde{Q}\left(j\right) &= \sum_{k=1}^{j}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathrm{Re}\left(I_{a\varepsilon,k}-A_{X_a,k}I_{\zeta,k}A_{\varepsilon,k}^*\right)\\ &= \sum_{k=1}^{l}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathrm{Re}\left(I_{a\varepsilon,k}-A_{X_a,k}I_{\zeta,k}A_{\varepsilon,k}^*\right)+\sum_{k=l+1}^{j}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathrm{Re}\left(I_{a\varepsilon,k}-A_{X_a,k}I_{\zeta,k}A_{\varepsilon,k}^*\right)\\ &\equiv Q_1+Q_2,\end{aligned}$$

with l being an integer less than j that is determined later. Firstly for Q_1, for each $1\leq k\leq l$ we consider $q_{1k}\equiv\mathrm{Re}\left(I_{a\varepsilon,k}-A_{X_a,k}I_{\zeta,k}A_{\varepsilon,k}^*\right)$,

$$\mathbb{E}\left|Q_1\right|\leq\sum_{k=1}^{l}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathbb{E}\left|q_{1k}\right|=\sum_{k=1}^{l}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathbb{E}\left|\mathrm{Re}\left(w_{X_a,k}w_{\varepsilon,k}^*-A_{X_a,k}w_{\zeta,k}w_{\zeta,k}^*A_{\varepsilon,k}^*\right)\right|$$

$$\leq \frac{1}{2}\sum_{k=1}^{l}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathbb{E}\left|\left(w_{X_a,k}-A_{X_a,k}w_{\zeta,k}\right)\left(w_{\varepsilon,k}^*+w_{\zeta,k}^*A_{\varepsilon,k}^*\right)\right|$$
$$+\frac{1}{2}\sum_{k=1}^{l}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\mathbb{E}\left|\left(w_{X_a,k}+A_{X_a,k}w_{\zeta,k}\right)\left(w_{\varepsilon,k}^*-w_{\zeta,k}^*A_{\varepsilon,k}^*\right)\right|$$
$$\leq \frac{1}{2}\left(\sum_{k=1}^{l}\lambda_k^{2d_{X_a}(u)}\mathbb{E}\left|w_{X_a,k}-A_{X_a,k}w_{\zeta,k}\right|^2\right)^{\frac{1}{2}}\left(\sum_{k=1}^{l}\lambda_k^{2d_\varepsilon(u)}\mathbb{E}\left|w_{\varepsilon,k}^*+w_{\zeta,k}^*A_{\varepsilon,k}^*\right|^2\right)^{\frac{1}{2}}$$
$$+\frac{1}{2}\left(\sum_{k=1}^{l}\lambda_k^{2d_{X_a}(u)}\mathbb{E}\left|w_{X_a,k}+A_{X_a,k}w_{\zeta,k}\right|^2\right)^{\frac{1}{2}}\left(\sum_{k=1}^{l}\lambda_k^{2d_\varepsilon(u)}\mathbb{E}\left|w_{\varepsilon,k}^*-w_{\zeta,k}^*A_{\varepsilon,k}^*\right|^2\right)^{\frac{1}{2}},$$

following the same reasoning in the proof of Lemma A.1.1. Then we have for $\mathbb{E}\left|\widetilde{w}_{\widetilde{X}_a,k}-A_{X_a,j}\widetilde{w}_{\widetilde{\zeta},k}\right|^2$ at each $k=1,\ldots,l$ that

$$\begin{aligned}
&\mathbb{E}\left|w_{X_a,k}-A_{X_a,k}w_{\zeta,k}\right|^2\\
&=\mathbb{E}\left[\left(w_{X_a,k}-A_{X_a,k}w_{\zeta,k}\right)\left(w_{X_a,k}^*-w_{\zeta,k}^*A_{X_a,k}^*\right)\right]\\
&=\frac{1}{c_{Tu}^2}\mathbb{E}\left[\sum_{t=1}^{T}\left(X_{a,t,T}-A_{X_a,k}\zeta_t\right)K_{h,tu}e^{it\lambda_k}\sum_{s=1}^{T}\left(X_{a,s,T}-\zeta_s'A_{X_a,k}^*\right)K_{h,su}e^{-is\lambda_k}\right]\\
&=\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\sum_{s=1}^{T}K_{h,tu}K_{h,su}e^{i(t-s)\lambda_k}\\
&\times\left[\mathbb{E}\left(X_{a,t,T}X_{a,s,T}\right)-\mathbb{E}\left(X_{a,t,T}\zeta_s'\right)A_{X_a,k}^*-A_{X_a,k}\mathbb{E}\left(\zeta_tX_{a,s,T}\right)+A_{X_a,k}\mathbb{E}\left(\zeta_t\zeta_s'\right)A_{X_a,k}^*\right]\\
&=\int_{-\pi}^{\pi}\widetilde{K}(\lambda)\left[A_{X_a}(u,\lambda+\lambda_k)-A_{X_a}(u,\lambda_k)\right]\left[A_{X_a}^*(u,\lambda+\lambda_k)-A_{X_a}^*(u,\lambda_k)\right]d\lambda\cdot O\left(1+\psi_1\right)\\
&\equiv E_{k1}+E_{k2}
\end{aligned}\tag{A.2.15}$$

where E_{k1} is the term containing the $O(1)$ multiplier and E_{k2} is the one with $O(\psi_1)$. The above is obtained as we have showed before in (A.2.9) in the proof of Proposition A.1.3, where $\psi_1=T^{-1}+\left|u-\frac{t}{T}\right|+\left|u-\frac{s}{T}\right|$, and $\widetilde{K}(\lambda)$ is given by

$$\widetilde{K}(\lambda)=\frac{1}{c_{Tu}^2}\left|\sum_{t=1}^{T}K_{h,tu}e^{it\lambda}\right|^2=O\left(T^{-1}h^{-1}|\lambda|^{-2}\right),$$

which is identical to the weighted Fejér kernel in (A.2.11) while we suppress its dependence on u. Note that E_{k2} is dominated by E_{k1} in order; and for E_{k1}, we have

$$\begin{aligned}
E_{k1}&\leq O(1)\int_{-\pi}^{\pi}\widetilde{K}(\lambda)\left[C_1f_{X_{aa}}(u,\lambda+\lambda_k)+C_2f_{X_{aa}}(u,\lambda_k)\right]d\lambda\\
&\leq C\int_{-\pi}^{\pi}\widetilde{K}(\lambda)\left[f_{X_{aa}}(u,\lambda+\lambda_k)+f_{X_{aa}}(u,\lambda_k)\right]d\lambda\equiv C\cdot E_{k1}^*
\end{aligned}$$

for some absolute constants C_1, C_2 and C. Then we decompose E_{k1}^* as

$$\begin{aligned} E_{k1}^* &= \left[\left| \int_{\varepsilon}^{\pi} \right| + \left| \int_{\lambda_k}^{\varepsilon} \right| + \int_{-\lambda_k}^{\lambda_k} \right] \widetilde{K}(\lambda) \left[f_{X_{aa}}(u, \lambda + \lambda_k) + f_{X_{aa}}(u, \lambda_k) \right] d\lambda \\ &\equiv E_{k11}^* + E_{k12}^* + E_{k13}^*, \end{aligned}$$

where ε is some fixed constant that is small enough. Firstly E_{k11}^* is further given by

$$\begin{aligned} E_{k11}^* &= \left(\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi} \right) \widetilde{K}(\lambda) \left[f_{X_{aa}}(u, \lambda + \lambda_k) + f_{X_{aa}}(u, \lambda_k) \right] d\lambda \\ &\leq \max_{|\lambda| \geq \varepsilon} \widetilde{K}(\lambda) \int_{-\pi}^{\pi} \left| f_{X_{aa}}(u, \lambda + \lambda_k) + f_{X_{aa}}(u, \lambda_k) \right| d\lambda \\ &= O\left(T^{-1} h^{-1} \left(1 + \lambda_k^{-2d_{X_a}(u)} \right) \right) = o\left(h^{-1} k^{-1} \lambda_k^{-2d_{X_a}(u)} \right), \end{aligned}$$

where the second equality holds by the nonnegativity of $f_{X_{aa}}(u, \lambda)$ and its integrability; and then E_{k12}^* follows that

$$\begin{aligned} E_{k12}^* &= \left(\int_{-\varepsilon}^{-\lambda_k} + \int_{\lambda_k}^{\varepsilon} \right) \widetilde{K}(\lambda) \left[f_{X_{aa}}(u, \lambda + \lambda_k) + f_{X_{aa}}(u, \lambda_k) \right] d\lambda \\ &\leq C \left[\max_{2\lambda_k \leq \lambda \leq \varepsilon + \lambda_k} \frac{|f_{X_{aa}}(u, \lambda)|}{\lambda^{(1-2d_{X_a}(u))/2}} \int_{2\lambda_k}^{\varepsilon + \lambda_k} \lambda^{(1-2d_{X_a}(u))/2} \widetilde{K}(\lambda) d\lambda + |f_{X_{aa}}(u, \lambda_k)| \int_{\lambda_k}^{\varepsilon} \widetilde{K}(\lambda) d\lambda \right] \\ &= O\left(\lambda_k^{-(1+2d_{X_a}(u))/2} \right) O\left(T^{-1} h^{-1} \int_{2\lambda_k}^{\infty} \lambda^{-(3+2d_{X_a}(u))/2} d\lambda \right) + O\left(\lambda_k^{-2d_{X_a}(u)} \right) O\left(T^{-1} h^{-1} \int_{\lambda_k}^{\infty} \lambda^{-2} d\lambda \right) \\ &= O\left(h^{-1} k^{-1} \lambda_k^{-2d_{X_a}(u)} \right) \end{aligned}$$

for some absolute constant C; next E_{k13}^* is given by

$$\begin{aligned} E_{k13}^* &= \int_{-\lambda_k}^{\lambda_k} \widetilde{K}(\lambda) \left[f_{X_{aa}}(u, \lambda + \lambda_k) + f_{X_{aa}}(u, \lambda_k) \right] d\lambda \\ &\leq 2 f_{X_{aa}}(u, \lambda_k) \int_{-\lambda_k}^{\lambda_k} \widetilde{K}(\lambda) d\lambda \\ &= O\left(h^{-1} k^{-1} \lambda_k^{-2d_{X_a}(u)} \right), \end{aligned}$$

therefore in conclusion $E_{k1}^* = O\left(h^{-1} k^{-1} \lambda_k^{-2d_{X_a}(u)} \right)$ for each $k = 1, \ldots, l$ and so as E_{k1} and $\mathbb{E} \left| w_{X_a,k} - A_{X_a,k} w_{\zeta,k} \right|^2$, and it is easy to see that the same order holds for $\mathbb{E} \left| w_{X_a,k} + A_{X_a,k} w_{\zeta,k} \right|^2$; and the same reasoning holds for $\mathbb{E} \left| w_{\varepsilon,k}^* - w_{\zeta,k}^* A_{\varepsilon,k}^* \right|^2$ and $\mathbb{E} \left| w_{\varepsilon,k}^* + w_{\zeta,k}^* A_{\varepsilon,k}^* \right|^2$. Therefore $Q_1 = O_p\left(h^{-1} \log l \right)$ by Riemann sum approximation.

Next for Q_2, we consider its expectation of squared norm following the proof of (C.2) in

Lobato (1999), which is given by

$$\begin{aligned}
\mathbb{E}\left|Q_2\right|^2 &= \mathbb{E}\left[\left(\sum_{k=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \left(I_{a\varepsilon,k} - A_{X_a,k} I_{\zeta,k} A^*_{\varepsilon,k}\right)\right)\left(\sum_{k=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \left(I^*_{a\varepsilon,k} - A_{\varepsilon,k} I_{\zeta,k} A^*_{X_a,k}\right)\right)\right] \\
&= \sum_{k=l+1}^{j} \lambda_k^{2(d_{X_a}(u)+d_\varepsilon(u))} \mathbb{E}\left[\left(w_{X_a,k} w^*_{\varepsilon,k} - A_{X_a,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{\varepsilon,k}\right)\left(w_{\varepsilon,k} w^*_{X_a,k} - A_{\varepsilon,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{X_a,k}\right)\right] \\
&+ \sum_{s\neq k; k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} \\
&\times \mathbb{E}\left[\left(w_{X_a,k} w^*_{\varepsilon,k} - A_{X_a,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{\varepsilon,k}\right)\left(w_{\varepsilon,s} w^*_{X_a,s} - A_{\varepsilon,s} w_{\zeta,s} w^*_{\zeta,s} A^*_{X_a,s}\right)\right] \\
&\equiv Q_{21} + Q_{22},
\end{aligned}$$

where in detail,

$$\begin{aligned}
Q_{21} = \sum_{k=l+1}^{j} \lambda_k^{2(d_{X_a}(u)+d_\varepsilon(u))} &\left[\mathbb{E}\left(w_{X_a,k} w^*_{\varepsilon,k} w_{\varepsilon,k} w^*_{X_a,k}\right) - \mathbb{E}\left(A_{X_a,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{\varepsilon,k} w_{\varepsilon,k} w^*_{X_a,k}\right)\right. \\
&\left. -\mathbb{E}\left(w_{X_a,k} w^*_{\varepsilon,k} A_{\varepsilon,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{X_a,k}\right) + \mathbb{E}\left(A_{X_a,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{\varepsilon,k} A_{\varepsilon,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{X_a,k}\right)\right]
\end{aligned}$$

and

$$\begin{aligned}
Q_{22} = \sum_{s\neq k; i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} &\left[\mathbb{E}\left(w_{X_a,k} w^*_{\varepsilon,k} w_{\varepsilon,s} w^*_{X_a,s}\right) - \mathbb{E}\left(A_{X_a,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{\varepsilon,k} w_{\varepsilon,s} w^*_{X_a,s}\right)\right. \\
&\left. -\mathbb{E}\left(w_{X_a,k} w^*_{\varepsilon,k} A_{\varepsilon,s} w_{\zeta,s} w^*_{\zeta,s} A^*_{X_a,s}\right) + \mathbb{E}\left(A_{X_a,k} w_{\zeta,k} w^*_{\zeta,k} A^*_{\varepsilon,k} A_{\varepsilon,s} w_{\zeta,s} w^*_{\zeta,s} A^*_{X_a,s}\right)\right].
\end{aligned}$$

Note that $w_{X_a,k}$, $w_{\varepsilon,k}$, $A_{X_a,k} w_{\zeta,k}$ and $A_{\varepsilon,k} w_{\zeta,k}$ are all mean-zero scalars for $k = l+1, \ldots, j$, and thus both Q_{21} and Q_{22} consist of terms in the form $\mathbb{E}(wxyz)$ with w, x, y and z as zero mean scalars, which follows that

$$\mathbb{E}(wxyz) = \mathbb{E}(wx)\,\mathbb{E}(yz) + \mathbb{E}(wy)\,\mathbb{E}(xz) + \mathbb{E}(wz)\,\mathbb{E}(xy) + \mathrm{cum}(w, x, y, z)$$

where $\mathrm{cum}(w, x, y, z)$ is the joint cumulant of these four random variables. For ease of notation, we denote $w_{X_a,k} \equiv a_k$, $w_{\varepsilon,k} \equiv e_k$, $A_{X_a,k} w_{\zeta,k} \equiv A_k$ and $A_{\varepsilon,k} w_{\zeta,k} \equiv E_k$. Then by decomposing Q_{21} and Q_{22} into two parts, with and without cumulants, we have them given by

$$\begin{aligned}
Q_{21} = \sum_{k=l+1}^{j} \lambda_k^{2(d_{X_a}(u)+d_\varepsilon(u))} &[\mathbb{E}(a_k e^*_k)\,\mathbb{E}(e_k a^*_k) + \mathbb{E}(a_k e_k)\,\mathbb{E}(e^*_k a^*_k) + \mathbb{E}(a_k a^*_k)\,\mathbb{E}(e^*_k e_k) \\
&- \mathbb{E}(A_k E^*_k)\,\mathbb{E}(e_k a^*_k) - \mathbb{E}(A_k e_k)\,\mathbb{E}(E^*_k a^*_k) - \mathbb{E}(A_k a^*_k)\,\mathbb{E}(E^*_k e_k) \\
&- \mathbb{E}(a_k e^*_k)\,\mathbb{E}(E_k A^*_k) - \mathbb{E}(a_k E_k)\,\mathbb{E}(e^*_k A^*_k) - \mathbb{E}(a_k A^*_k)\,\mathbb{E}(e^*_k E_k)
\end{aligned}$$

$$+\mathbb{E}(A_kE_k^*)\mathbb{E}(E_kA_k^*)+\mathbb{E}(A_kE_k)\mathbb{E}(E_k^*A_k^*)+\mathbb{E}(A_kA_k^*)\mathbb{E}(E_k^*E_k)]$$

$$+\sum_{k=l+1}^{j}\lambda_k^{2(d_{X_a}(u)+d_\varepsilon(u))}[\mathrm{cum}(a_k,e_k^*,e_k,a_k^*)-\mathrm{cum}(A_k,E_k^*,e_k,a_k^*)$$

$$-\mathrm{cum}(a_k,e_k^*,E_k,A_k^*)+\mathrm{cum}(A_k,E_k^*,E_k,A_k^*)]\equiv Q_{21,a}+Q_{21,b},$$

where $Q_{21,a}$ is the part without cumulant and $Q_{21,b}$ is the part with cumulant, and using the same notataion,

$$Q_{22}=\sum_{s\neq k;i,k,s=l+1}^{j}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\lambda_s^{d_{X_a}(u)+d_\varepsilon(u)}[\mathbb{E}(a_ke_k^*)\mathbb{E}(e_sa_s^*)+\mathbb{E}(a_ke_s)\mathbb{E}(e_k^*a_s^*)+\mathbb{E}(a_ka_s^*)\mathbb{E}(e_k^*e_s)$$

$$-\mathbb{E}(A_kE_k^*)\mathbb{E}(e_sa_s^*)-\mathbb{E}(A_ke_s)\mathbb{E}(E_k^*a_s^*)-\mathbb{E}(A_ka_s^*)\mathbb{E}(E_k^*e_s)$$

$$-\mathbb{E}(a_ke_k^*)\mathbb{E}(E_sA_s^*)-\mathbb{E}(a_kE_s)\mathbb{E}(e_k^*A_s^*)-\mathbb{E}(a_kA_s^*)\mathbb{E}(e_k^*E_s)$$

$$+\mathbb{E}(A_kE_k^*)\mathbb{E}(E_sA_s^*)+\mathbb{E}(A_kE_s)\mathbb{E}(E_k^*A_s^*)+\mathbb{E}(A_kA_s^*)\mathbb{E}(E_k^*E_s)]$$

$$+\sum_{s\neq k;i,k,s=l+1}^{j}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\lambda_s^{d_{X_a}(u)+d_\varepsilon(u)}$$

$$\times[\mathrm{cum}(a_k,e_k^*,e_s,a_s^*)-\mathrm{cum}(A_k,E_k^*,e_s,a_s^*)-\mathrm{cum}(a_k,e_k^*,E_s,A_s^*)+\mathrm{cum}(A_k,E_k^*,E_s,A_s^*)]$$

$$\equiv Q_{22,a}+Q_{22,b},$$

where $Q_{22,a}$ is the part without cumulant and $Q_{22,b}$ is the part with cumulant. Firstly we derive the order for $Q_{21,a}$ and $Q_{22,a}$. Since we need to obtain some similar results as Theorem 2 in Robinson (1995c), we consider analyzing $\mathbb{E}(a_ke_k^*)$ as the following:

$$\mathbb{E}(a_ke_k^*)=\mathbb{E}\left(w_{X_a,k}w_{\varepsilon,k}^*\right)=\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(X_{a,t,T}\varepsilon_{s,T})K_{h,tu}K_{h,su}e^{i(t-s)\lambda_k}$$

$$=\frac{1}{c_{Tu}^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\int_{-\pi}^{\pi}f_{X_a\varepsilon}(u,\lambda)e^{-i(t-s)\lambda}d\lambda\cdot O(1+\psi_1)\right]K_{h,tu}K_{h,su}e^{i(t-s)\lambda_k}$$

$$=\int_{-\pi}^{\pi}f_{X_a\varepsilon}(u,\lambda)\widetilde{K}(\lambda_k-\lambda)d\lambda\cdot O(1+h)$$

$$=\int_{-\pi}^{\pi}(f_{X_a\varepsilon}(u,\lambda)-f_{X_a\varepsilon}(u,\lambda_k))\widetilde{K}(\lambda_k-\lambda)d\lambda\cdot O(1+h)+f_{X_a\varepsilon}(u,\lambda_k)\cdot O(1+h)$$

$$\equiv EW_1+EW_2$$

where $EW_1=O(1)\int_{-\pi}^{\pi}(f_{X_a\varepsilon}(u,\lambda)-f_{X_a\varepsilon}(u,\lambda_k))\widetilde{K}(\lambda_k-\lambda)d\lambda$, and EW_2 refers to the rest of the terms. Next we denote C as some absolute constant whose value may varies across lines. Following the proof of Theorem 2 in Robinson (1995c),

$$EW_1\leq C\int_{-\pi}^{\pi}(f_{X_a\varepsilon}(u,\lambda)-f_{X_a\varepsilon}(u,\lambda_k))\widetilde{K}(\lambda_k-\lambda)d\lambda$$

$$= C\left[\left|\int_{\varepsilon}^{\pi}\right| + \left|\int_{2\lambda_k}^{\varepsilon}\right| + \left|\int_{\lambda_k/2}^{2\lambda_k}\right| + \int_{-\lambda_k/2}^{\lambda_k/2}\right]\left(f_{X_a\varepsilon}(u,\lambda) - f_{X_a\varepsilon}(u,\lambda_k)\right)\widetilde{K}(\lambda_k - \lambda)\,d\lambda$$

$$\equiv C\left(EW_{11} + EW_{12} + EW_{13} + EW_{14}\right),$$

where firstly EW_{11} is given by

$$\begin{aligned}
EW_{11} &= \left(\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi}\right)\left(f_{X_a\varepsilon}(u,\lambda) - f_{X_a\varepsilon}(u,\lambda_k)\right)\widetilde{K}(\lambda_k - \lambda)\,d\lambda \\
&\leq C\max_{|\lambda|\geq\varepsilon}\widetilde{K}(\lambda_k - \lambda)\int_{-\pi}^{\pi}\left|f_{X_a\varepsilon}(u,\lambda) - f_{X_a\varepsilon}(u,\lambda_k)\right|d\lambda \\
&\leq C\max_{|\lambda|\geq\varepsilon}\widetilde{K}(\lambda_k - \lambda)\int_{-\pi}^{\pi}\left|f_{X_a\varepsilon}(u,\lambda)\right| + \left|f_{X_a\varepsilon}(u,\lambda_k)\right|d\lambda \\
&\leq \frac{C}{Th}\left(1 + \lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}\right).
\end{aligned}$$

Note that the above order holds by Assumption 2(i), $\widetilde{K}(\lambda) \leq CT^{-1}h^{-1}\lambda^{-2}$, and the fact that $f_{X_a\varepsilon}(u,\lambda) \geq 0$ and $\int_{-\pi}^{\pi} f_{X_a\varepsilon}(u,\lambda)\,d\lambda < \infty$. Next for EW_{12} we have

$$\begin{aligned}
&EW_{12} \\
&= \left(\int_{-\varepsilon}^{-2\lambda_k} + \int_{2\lambda_k}^{\varepsilon}\right)\left(f_{X_a\varepsilon}(u,\lambda) - f_{X_a\varepsilon}(u,\lambda_k)\right)\widetilde{K}(\lambda_k - \lambda)\,d\lambda \\
&\leq C\max_{2\lambda_k\leq\lambda\leq\varepsilon}\frac{f_{X_a\varepsilon}(u,\lambda)}{\lambda^{(1-d_{X_a}(u)-d_\varepsilon(u))/2}}\int_{2\lambda_k}^{\pi}\lambda^{(1-d_{X_a}(u)-d_\varepsilon(u))/2}\widetilde{K}(\lambda_k - \lambda)\,d\lambda + \left|f_{X_a\varepsilon}(u,\lambda_k)\right|\int_{2\lambda_k}^{\pi}\widetilde{K}(\lambda_k - \lambda)\,d\lambda \\
&= C\max_{2\lambda_k\leq\lambda\leq\varepsilon}\lambda^{-(1+d_{X_a}(u)+d_\varepsilon(u))/2}\frac{1}{Th}\int_{2\lambda_k}^{\pi}\lambda^{-(3+d_{X_a}(u)+d_\varepsilon(u))/2}d\lambda + C\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}\frac{1}{Th}\int_{2\lambda_k}^{\pi}\lambda^{-2}d\lambda \\
&= Ch^{-1}k^{-1}\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)},
\end{aligned}$$

using the similar reasoing as before. Next for EW_{13} we have

$$\begin{aligned}
EW_{13} &= \left(\int_{-2\lambda_k}^{-\lambda_k/2} + \int_{\lambda_k/2}^{2\lambda_k}\right)\left(f_{X_a\varepsilon}(u,\lambda) - f_{X_a\varepsilon}(u,\lambda_k)\right)\widetilde{K}(\lambda_k - \lambda)\,d\lambda \\
&\leq C\max_{\lambda_k/2\leq\lambda\leq2\lambda_k}\left|\frac{\partial f_{X_a\varepsilon}(u,\lambda)}{\partial\lambda}\right|\int_{\lambda_k/2}^{2\lambda_k}\left|\lambda_k - \lambda\right|\widetilde{K}(\lambda_k - \lambda)\,d\lambda \\
&\leq C\lambda_k^{-1-d_{X_a}(u)-d_\varepsilon(u)}\frac{1}{T}\int_{\lambda_k/2}^{2\lambda_k}\left|\sum_{t=1}^{T}K_{h,tu}e^{it(\lambda_k-\lambda)}\right|d\lambda,
\end{aligned}$$

and it remains to derive the order for $\int_0^{C\lambda_k}\left|\sum_{t=1}^{T}K_{h,tu}e^{it\lambda}\right|d\lambda$ for some constant C by change of variable. By our proof of Proposition A.1.3 we have $\left|\sum_{t=1}^{T}K_{h,tu}e^{it\lambda}\right| \leq Ch^{-1}\left|\frac{\sin(T\lambda/2)}{\sin(\lambda/2)}\right|$ for some constant C, and thus

$$\int_0^{C\lambda_k}\left|\sum_{t=1}^{T}K_{h,tu}e^{it\lambda}\right|d\lambda \leq Ch^{-1}\int_0^{C\lambda_k}\left|\frac{\sin(T\lambda/2)}{\sin(\lambda/2)}\right|d\lambda \leq Ch^{-1}\log k$$

using Lemma 5 in Robinson (1994a) for some constant C, when k is large enough but fixed.[1] Then we can conclude that $EW_{13} \leq Ch^{-1}k^{-1}\log k\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}$. At last for EW_{14} we have

$$
\begin{aligned}
EW_{14} &= \int_{-\lambda_k/2}^{\lambda_k/2} \left(f_{X_a\varepsilon}(u,\lambda) - f_{X_a\varepsilon}(u,\lambda_k)\right)\widetilde{K}(\lambda_k-\lambda)\,d\lambda \\
&\leq \max_{|\lambda|\leq\lambda_k/2}\widetilde{K}(\lambda_k-\lambda)\int_{-\lambda_k/2}^{\lambda_k/2}|f_{X_a\varepsilon}(u,\lambda)| + |f_{X_a\varepsilon}(u,\lambda_k)|\,d\lambda \\
&\leq \frac{C}{Th}\lambda_k^{-2}\lambda_k^{1-d_{X_a}(u)-d_\varepsilon(u)} \leq Ch^{-1}k^{-1}\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}
\end{aligned}
$$

for some absolute constant C. Note that $h^{-1}k^{-1}\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}$ is dominated by $h^{-1}k^{-1}\log k\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}$, we thus have $EW_1 \leq Ch^{-1}k^{-1}\log k\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}$, and thus EW_2 is dominated by EW_1, which then concludes that

$$
\mathbb{E}\left(a_k e_k^*\right) = \mathbb{E}\left(w_{X_a,k}w_{\varepsilon,k}^*\right) \leq Ch^{-1}k^{-1}\log k\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}.
$$

And it is easy to see that $\mathbb{E}\left(e_k a_k^*\right)$ has the same order. Next using the same reasoning, we have

$$
\mathbb{E}\left(a_k a_k^*\right) = \mathbb{E}\left(w_{X_a,k}w_{X_a,k}^*\right) \leq Ch^{-1}k^{-1}\log k\lambda_k^{-2d_{X_a}(u)},
$$

$$
\mathbb{E}\left(e_k e_k^*\right) = \mathbb{E}\left(w_{\varepsilon,k}w_{\varepsilon,k}^*\right) \leq Ch^{-1}k^{-1}\log k\lambda_k^{-2d_\varepsilon(u)}.
$$

And it is easy to see that

$$
\begin{aligned}
\mathbb{E}\left(A_k E_k^*\right) &= \mathbb{E}\left(A_{X_a,k}w_{\zeta,k}w_{\zeta,k}^*A_{\varepsilon,k}^*\right) = A_{X_a,k}\frac{1}{c_{Tu}^2}\sum_{t,s=1}^{T}\mathbb{E}\left(\zeta_t\zeta_s^{'}\right)K_{h,tu}K_{h,su}e^{i(t-s)\lambda_k}A_{\varepsilon,k}^* \\
&= \frac{1}{2\pi}A_{X_a,k}A_{\varepsilon,k}^* = f_{X_a\varepsilon}(u,\lambda_k) \leq C\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}
\end{aligned}
$$

by Assumption 1(i) and 2(i). And it is easy to see the same bound holds for $\mathbb{E}\left(E_k A_k^*\right)$, and by the same reasoning and the fact that E_k is a scalar,

$$
\mathbb{E}\left(E_k^* E_k\right) = \mathbb{E}\left(E_k E_k^*\right) \leq C\lambda_k^{-2d_\varepsilon(u)} \text{ and } \mathbb{E}\left(A_k A_k^*\right) \leq C\lambda_k^{-2d_{X_a}(u)}.
$$

Next using Cauchy-Schwarz inequality, $\mathbb{E}\left(A_k a_k^*\right)$ can be bounded by

$$
\left|\mathbb{E}\left(A_k a_k^*\right)\right| \leq \left(\mathbb{E}\left(A_k A_k^*\right)\right)^{\frac{1}{2}}\left(\mathbb{E}\left(a_k a_k^*\right)\right)^{\frac{1}{2}} \leq C\sqrt{h^{-1}k^{-1}\log k}\lambda_k^{-2d_{X_a}(u)},
$$

[1] Note that in Lemma 5 of Robinson (1994a), k is required to be a sequence $k(T)$ such that $\frac{k}{T}\to 0$ as $T\to\infty$. But by its proof the order also holds when k is large enough but fixed. However even if we treat k as such a sequence, the same conclusion should also hold as in (A.2.19) the asymptotic order is determined by the sum of tail terms when j and k are large enough as sequences defined above.

and by the same reasoning we have

$$|\mathbb{E}(A_k e_k)| \leq C\sqrt{h^{-1}k^{-1}\log k}\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)},$$

$$|\mathbb{E}(a_k E_k)| \leq C\sqrt{h^{-1}k^{-1}\log k}\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)},$$

$$|\mathbb{E}(E_k^* e_k)| \leq C\sqrt{h^{-1}k^{-1}\log k}\lambda_k^{-2d_\varepsilon(u)}.$$

Then combining all the terms we analyzed so far, we can conclude that

$$\begin{aligned} Q_{21,a} = \sum_{k=l+1}^{j} \lambda_k^{2(d_{X_a}(u)+d_\varepsilon(u))} & \mathbb{E}(a_k e_k^*)\,\mathbb{E}(e_k a_k^*) + \mathbb{E}(a_k e_k)\,\mathbb{E}(e_k^* a_k^*) + \mathbb{E}(a_k a_k^*)\,\mathbb{E}(e_k^* e_k) \qquad \text{(A.2.16)}\\ & - \mathbb{E}(A_k E_k^*)\,\mathbb{E}(e_k a_k^*) - \mathbb{E}(A_k e_k)\,\mathbb{E}(E_k^* a_k^*) - \mathbb{E}(A_k a_k^*)\,\mathbb{E}(E_k^* e_k)\\ & - \mathbb{E}(a_k e_k^*)\,\mathbb{E}(E_k A_k^*) - \mathbb{E}(a_k E_k)\,\mathbb{E}(e_k^* A_k^*) - \mathbb{E}(a_k A_k^*)\,\mathbb{E}(e_k^* E_k)\\ & + \mathbb{E}(A_k E_k^*)\,\mathbb{E}(E_k A_k^*) + \mathbb{E}(A_k E_k)\,\mathbb{E}(E_k^* A_k^*) + \mathbb{E}(A_k A_k^*)\,\mathbb{E}(E_k^* E_k)\\ & = O\left(\sum_{k=l+1}^{j}\left(h^{-1}k^{-1}\log k\right)^2\right) = O\left(h^{-2}l^{-1}(\log j)^2\right). \end{aligned}$$

And using Cauchy-Schwarz inequality and the same reasoning above, we have for $Q_{22,a}$ that

$$\begin{aligned} Q_{22,a} = \sum_{s\neq k;i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} & \mathbb{E}(a_k e_k^*)\,\mathbb{E}(e_s a_s^*) + \mathbb{E}(a_k e_s)\,\mathbb{E}(e_k^* a_s^*) + \mathbb{E}(a_k a_s^*)\,\mathbb{E}(e_k^* e_s)\\ & - \mathbb{E}(A_k E_k^*)\,\mathbb{E}(e_s a_s^*) - \mathbb{E}(A_k e_s)\,\mathbb{E}(E_k^* a_s^*) - \mathbb{E}(A_k a_s^*)\,\mathbb{E}(E_k^* e_s)\\ & - \mathbb{E}(a_k e_k^*)\,\mathbb{E}(E_s A_s^*) - \mathbb{E}(a_k E_s)\,\mathbb{E}(e_k^* A_s^*) - \mathbb{E}(a_k A_s^*)\,\mathbb{E}(e_k^* E_s)\\ & + \mathbb{E}(A_k E_k^*)\,\mathbb{E}(E_s A_s^*) + \mathbb{E}(A_k E_s)\,\mathbb{E}(E_k^* A_s^*) + \mathbb{E}(A_k A_s^*)\,\mathbb{E}(E_k^* E_s)\\ & = O\left(\left(\sum_{k=l+1}^{j} h^{-1}k^{-1}\log k\right)^2\right) = O\left(h^{-2}(\log j)^4\right). \end{aligned}$$

Therefore by specifying $l \sim j^\alpha$ for an arbitrary $\alpha \in (0,1)$, we can conclude that

$$Q_{21,a} + Q_{22,a} = O\left(h^{-2}j^{-\alpha}(\log j)^2 + h^{-2}(\log j)^4\right). \qquad \text{(A.2.17)}$$

Next for the parts containing cumulants, $Q_{21,b}$ and $Q_{22,b}$, following (C.9) in Lobato (1999), it is sufficient to analyze $Q_{22,b}$ as it is the dominant one, which is given by

$$\begin{aligned} Q_{22,b} = & \sum_{s\neq k;i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\lambda_s^{d_{X_a}(u)+d_\varepsilon(u)}\\ & \times \left[\operatorname{cum}(a_k, e_k^*, e_s, a_s^*) - \operatorname{cum}(A_k, E_k^*, e_s, a_s^*) - \operatorname{cum}(a_k, e_k^*, E_s, A_s^*) + \operatorname{cum}(A_k, E_k^*, E_s, A_s^*)\right] \end{aligned}$$

$$= \sum_{s\neq k;i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} \left[\mathrm{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon},k}, \widetilde{w}_{\widetilde{\varepsilon},s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right)\right.$$
$$- \mathrm{cum}\left(A_{X_a,k}\widetilde{w}_{\widetilde{\zeta},k}, \widetilde{w}^*_{\widetilde{\zeta},k}A^*_{\varepsilon,k}, \widetilde{w}_{\widetilde{\varepsilon},s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right)$$
$$\left.-\mathrm{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon},k}, A_{\varepsilon,s}\widetilde{w}_{\widetilde{\zeta},s}, \widetilde{w}^*_{\widetilde{\zeta},s}A_{X_a,s}\right) + \mathrm{cum}\left(A_{X_a,k}\widetilde{w}_{\widetilde{\zeta},k}, \widetilde{w}^*_{\widetilde{\zeta},k}A^*_{\varepsilon,k}, A_{\varepsilon,s}\widetilde{w}_{\widetilde{\zeta},s}, \widetilde{w}^*_{\widetilde{\zeta},s}A_{X_a,s}\right)\right],$$

where we denote

$$\widetilde{w}_{\widetilde{X}_a,k} = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \widetilde{X}_{a,t,u,T} e^{it\lambda_k} = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \sqrt{\frac{2\pi T}{c_{Tu}^2}} X_{a,t,T} K_{h,tu} e^{it\lambda_k},$$

and denote $\widetilde{w}^*_{\widetilde{\varepsilon},k}$ and $\widetilde{w}_{\widetilde{\zeta},k}$ in the same way. We firstly consider $\mathrm{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon},k}, \widetilde{w}_{\widetilde{\varepsilon},s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right)$, which is given by

$$\mathrm{cum}\left(\widetilde{w}_{\widetilde{X}_a,k}, \widetilde{w}^*_{\widetilde{\varepsilon},k}, \widetilde{w}_{\widetilde{\varepsilon},s}, \widetilde{w}^*_{\widetilde{X}_a,s}\right) \tag{A.2.18}$$
$$= \mathrm{cum}\left(\frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \widetilde{X}_{a,t,u,T} e^{it\lambda_k}, \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \widetilde{\varepsilon}_{t,u,T} e^{-it\lambda_k},\right.$$
$$\left.\frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \widetilde{\varepsilon}_{t,u,T} e^{it\lambda_s}, \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \widetilde{X}_{a,t,u,T} e^{-it\lambda_s}\right)$$
$$= \frac{\sum_{t_1,\ldots,t_4=1}^{T}}{(2\pi T)^2} \mathrm{cum}\left(\widetilde{X}_{a,t_1,u,T}, \widetilde{\varepsilon}_{t_2,u,T}, \widetilde{\varepsilon}_{t_3,u,T}, \widetilde{X}_{a,t_4,u,T}\right) e^{i[(t_1-t_2)\lambda_k+(t_3-t_4)\lambda_s]}$$
$$= \frac{\sum_{t_1,\ldots,t_4=1}^{T}}{(2\pi T)^2} \mathrm{cum}\left(\sum_{j=0}^{\infty} A_{X_a,t_1,T}(j)\,\zeta_{t_1-j,T}, \sum_{j=0}^{\infty} A_{\varepsilon,t_2,T}(j)\,\zeta_{t_2-j,T},\right.$$
$$\left.\sum_{j=0}^{\infty} A_{\varepsilon,t_3,T}(j)\,\zeta_{t_3-j,T}, \sum_{j=0}^{\infty} A_{X_a,t_4,T}(j)\,\zeta_{t_4-j,T}\right)$$
$$\times \left(\frac{2\pi T}{c_{Tu}^2}\right)^2 K_{h,t_1u}K_{h,t_2u}K_{h,t_3u}K_{h,t_4u} e^{i[(t_1-t_2)\lambda_k+(t_3-t_4)\lambda_s]}$$
$$= \frac{\sum_{t_1,\ldots,t_4=1}^{T}}{(2\pi T)^2} \mathrm{cum}\left(\sum_{j=0}^{\infty} A^0_{X_a,j}\zeta_{t_1-j,T}, \sum_{j=0}^{\infty} A^0_{\varepsilon,j}\zeta_{t_2-j,T}, \sum_{j=0}^{\infty} A^0_{\varepsilon,j}\zeta_{t_3-j,T}, \sum_{j=0}^{\infty} A^0_{X_a,j}\zeta_{t_4-j,T}\right)$$
$$\times \left(\frac{2\pi T}{c_{Tu}^2}\right)^2 K_{h,t_1u}K_{h,t_2u}K_{h,t_3u}K_{h,t_4u} e^{i[(t_1-t_2)\lambda_k+(t_3-t_4)\lambda_s]} O(1+h)$$
$$\equiv \frac{\sum_{t_1,\ldots,t_4=1}^{T}}{(2\pi T)^2} \mathrm{cum}\left(G_1(t_1), G_2(t_2), G_3(t_3), G_4(t_4)\right) \left(\frac{2\pi T}{c_{Tu}^2}\right)^2 K_{h,t_{1-4}u} e^{i[(t_1-t_2)\lambda_k+(t_3-t_4)\lambda_s]} O(1+h)$$

by linearity of cumulants and Assumption 1, where we denote $A_{X_a,t_1,T}(j)$ as the a-th row of $A_{X,t_1,T}(j)$ and $A^0_{X_a,j} = A^0_{X_a}(u,j)$ and $A^0_{\varepsilon,j} = A^0_{\varepsilon}(u,j)$ using the notation in Assumption 1(ii). In the following we further define the vector

$$G(t) = (G_1(t), G_2(t), G_3(t), G_4(t))',$$

then $G(t)$ is a 4×1 vector-valued linear process given by

$$G_i(t) = \sum_{j=0}^{\infty} A^0_{G_\iota,j} \zeta_{t-j,T},\ \iota = 1, \ldots 4,$$

where

$$A^0_{G,j} = \left(A^{0\prime}_{G_1,j}, A^{0\prime}_{G_2,j}, A^{0\prime}_{G_3,j}, A^{0\prime}_{G_4,j}\right)' = \left(A^{0\prime}_{X_a,j}, A^{0\prime}_{\varepsilon,j}, A^{0\prime}_{\varepsilon,j}, A^{0\prime}_{X_a,j}\right)'$$

is the $4 \times (p+1)$ matrix of filter for $G(t)$. Note that the cumulants provided by (A.2.18) is derived in the same way as in Lobato (1999, pp. 146), except we involve the kernels $\left(\frac{2\pi T}{c^2_{Tu}}\right)^2 K_{h,t_1u} K_{h,t_2u} K_{h,t_3u} K_{h,t_4u}$ over summation of t. And it is the same for other cumumlants in $Q_{22,b}$. Therefore (C.9) in Lobato (1999, pp. 146) shall still hold with Dirichlet's kernel replaced by a kernelized version, as in the following:

$$\begin{aligned} Q_{22,b} = & \sum_{s\neq k; i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} \frac{1}{(2\pi)^3} \sum_{p_1,\ldots,p_4=1}^{p+1} \frac{\kappa_{p_1,\ldots p_4}}{(2\pi T)^2} \\ & \times \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left[A_{G_1p_1}(\lambda_s+\lambda_1+\lambda_2+\lambda_3)\overline{A_{G_2p_2}(\lambda_s-\lambda_2)} - A_{G_1p_1}(\lambda_s)\overline{A_{G_2p_2}(\lambda_s)}\right] \\ & \times \left[A_{G_3p_3}(\lambda_k-\lambda_1) A_{G_4p_4}(\lambda_k-\lambda_3) - A_{G_3p_3}(\lambda_k)\overline{A_{G_4p_4}(\lambda_k)}\right] \Xi_{ks}(\lambda_1,\lambda_2,\lambda_3)\, d\lambda_1 d\lambda_2 d\lambda_3 O(1+h), \end{aligned}$$

where $A_{G_1p_1}(\lambda) = \sum_{j=-\infty}^{\infty} A^0_{G_\iota,j} e^{ij\lambda}$, and $\kappa_{p_1,\ldots p_4} = \mathbb{E}(\zeta_{p_1}\zeta_{p_2}\zeta_{p3}\zeta_{p_4}) - 3$, and the same notation holds for other similar terms. And

$$\Xi_{ks}(\lambda_1,\lambda_2,\lambda_3) = \widetilde{D}(\lambda_s-\lambda_1-\lambda_2-\lambda_3)\,\widetilde{D}(\lambda_1+\lambda_k)\,\widetilde{D}(\lambda_2-\lambda_s)\,\widetilde{D}(\lambda_3-\lambda_k)$$

where $\widetilde{D}(\lambda) = \frac{1}{\sqrt{\frac{1}{2\pi T}c^2_{Tu}}} \sum_{t=1}^{T} K_{h,tu} e^{it\lambda}$ is the "kernelized" Dirichlet's kernel. Then following the reasoning in Lobato (1999, pp. 146-148), it is adequate to consider the order of

$$\widetilde{P}_j(I,p_k) = \int_{-\pi}^{\pi} \left|A_{Ip_k}(\lambda+\lambda_j) - A_{Ip_k}(\lambda_j)\right|^2 \widetilde{K}(\lambda-\lambda_j)\, d\lambda,\ I = X_a, \varepsilon,$$

where

$$\widetilde{K}(\lambda) = \frac{1}{c^2_{Tu}} \left|\sum_{t=1}^{T} K_{h,tu} e^{it\lambda}\right|^2 = \frac{1}{2\pi T}\left|\widetilde{D}(\lambda)\right|^2$$

because it is shown by Lobato (1999) that

$$\begin{aligned} Q_{22,b} = & O\left(\sum_{s\neq k; i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} \widetilde{P}_s^{\frac{1}{2}}(X_a,p_1)\,\widetilde{P}_s^{\frac{1}{2}}(\varepsilon,p_2)\,\widetilde{P}_k^{\frac{1}{2}}(X_a,p_3)\,\widetilde{P}_k^{\frac{1}{2}}(\varepsilon,p_4)\right) \\ & + O\left(\sum_{s\neq k; i,k,s=l+1}^{j} \lambda_k^{d_{X_a}(u)+d_\varepsilon(u)} \lambda_s^{d_{X_a}(u)+d_\varepsilon(u)} \widetilde{P}_s^{\frac{1}{2}}(X_a,p_1)\,\widetilde{P}_s^{\frac{1}{2}}(\varepsilon,p_2)\,\widetilde{P}_k^{\frac{1}{2}}(\varepsilon,p_3)\,\lambda_k^{-d_{X_a}(u)}\right) \end{aligned}$$

$$+O\left(\sum_{s\neq k;i,k,s=l+1}^{j}\lambda_k^{d_{X_a}(u)+d_\varepsilon(u)}\lambda_s^{d_{X_a}(u)+d_\varepsilon(u)}T^{-\frac{1}{2}}\widetilde{P}_s^{\frac{1}{2}}\left(X_a,p_1\right)\widetilde{P}_s^{\frac{1}{2}}\left(\varepsilon,p_2\right)\lambda_k^{-d_{X_a}(u)-d_\varepsilon(u)}\right).$$

And note that $\widetilde{P}_j\left(I,p_k\right)$ shares the same order for all p_k with (A.2.15) as we can replace λ with $\lambda+\lambda_j$ because of the periodicity of 2π for both $A_{Ip_k}(\cdot)$ and $\widetilde{K}(\cdot)$. Thus $\widetilde{P}_j\left(I,p_k\right)=O\left(h^{-1}j^{-1}\lambda_j^{-2d_I(u)}\right)$ uniformly for all for all p_k. Therefore by substitution

$$Q_{22,b}=O\left(h^{-2}\left(\log j\right)^2+h^{-2}j^{\frac{1}{2}}\log j+h^{-2}jT^{-\frac{1}{2}}\log j\right),$$

and together with (A.2.17), we can conclude that

$$Q_2=O\left(h^{-1}\left(\log j\right)^2+h^{-1}j^{\frac{1}{4}}\left(\log j\right)^{\frac{1}{2}}+h^{-1}j^{\frac{1}{2}}T^{-\frac{1}{4}}\left(\log j\right)^{\frac{1}{2}}\right)$$

and

$$\widetilde{Q}\left(j\right)=Q_1+Q_2=O_p\left(h^{-1}\left(\log j\right)^2+h^{-1}j^{\frac{1}{4}}\left(\log j\right)^{\frac{1}{2}}+h^{-1}j^{\frac{1}{2}}T^{-\frac{1}{4}}\left(\log j\right)^{\frac{1}{2}}\right)$$

by our choice of l. Then the order of $\widetilde{B}_{1,Ma}$ is given by

$$\begin{aligned}
&\widetilde{B}_{1,Ma}\\
&=\sum_{j=1}^{M-1}\left(\lambda_j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}-\lambda_{j+1}^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\right)\cdot\widetilde{Q}\left(j\right)+\lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\cdot\widetilde{Q}\left(M\right)\\
&=O_p\left(T^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sum_{j=1}^{M}j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)-1}\left(h^{-1}\left(\log j\right)^2+h^{-1}j^{\frac{1}{4}}\left(\log j\right)^{\frac{1}{2}}+h^{-1}j^{\frac{1}{2}}T^{-\frac{1}{4}}\left(\log j\right)^{\frac{1}{2}}\right)\right)\\
&+O_p\left(\lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\left(h^{-1}\left(\log M\right)^2+h^{-1}M^{\frac{1}{4}}\left(\log M\right)^{\frac{1}{2}}+h^{-1}M^{\frac{1}{2}}T^{-\frac{1}{4}}\left(\log M\right)^{\frac{1}{2}}\right)\right)\\
&=O_p\left(T^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\left(M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}h^{-1}\left(\log M\right)^2+M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)+\frac{1}{4}}h^{-1}\left(\log M\right)^{\frac{1}{2}}\right)\right)\\
&+O_p\left(M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)+\frac{1}{2}}h^{-1}T^{-\frac{1}{4}}\left(\log M\right)^{\frac{1}{2}}\right)\\
&+O_p\left(\lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\left(h^{-1}\left(\log M\right)^2+h^{-1}M^{\frac{1}{4}}\left(\log M\right)^{\frac{1}{2}}+h^{-1}M^{\frac{1}{2}}T^{-\frac{1}{4}}\left(\log M\right)^{\frac{1}{2}}\right)\right)\\
&=O_p\left(\lambda_M^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\left(h^{-1}\left(\log M\right)^2+h^{-1}M^{\frac{1}{4}}\left(\log M\right)^{\frac{1}{2}}+h^{-1}M^{\frac{1}{2}}T^{-\frac{1}{4}}\left(\log M\right)^{\frac{1}{2}}\right)\right),
\end{aligned}$$

which completes the proof of Proposition A.1.4.

Proof of Proposition A.1.5. (i) Recall that $Z_{t,T}(u)=\widetilde{\zeta}'_{t,u,T}\sum_{s<t}C_{t,s,T}(u)\widetilde{\zeta}_{s,u,T}$, where

$$\begin{aligned}
C_{t,T}(u)&=\sum_{a=1}^{p}\eta_a\frac{\lambda_M^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\sqrt{h}}{2\pi T\sqrt{M}}\sum_{j=1}^{M}\lambda_j^{2\delta(u)}\mathrm{Re}\left(A'_{X_a,j}\overline{A}_{\varepsilon,j}+A'_{\varepsilon,j}\overline{A}_{X_a,j}\right)\cos\left(t\lambda_j\right)\\
&\equiv\frac{\sqrt{h}}{2\pi T\sqrt{M}}\sum_{j=1}^{M}\theta_j\cos\left(t\lambda_j\right)
\end{aligned}$$

using the notation that $A_{X_a,j} = A_{X_a,j}(u,\lambda_j)$, $A_{\varepsilon,j} = A_{\varepsilon,j}(u,\lambda_j)$ and $\widetilde{\zeta}_{t,u,T} = \frac{\zeta_t K_{h,tu}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} K_{h,tu}^2}}$. Note that

$$
\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}\left(Z_{t,T}^2(u) \mid \mathcal{F}_{t-1}\right) \\
&= \sum_{t=2}^{T} \mathbb{E}\left(\sum_{s<t}\sum_{r<t} \widetilde{\zeta}_s' C_{t-s,T}'(u)\, \widetilde{\zeta}_t \widetilde{\zeta}_t' C_{t-r,T}(u)\, \widetilde{\zeta}_r \mid \mathcal{F}_{t-1}\right) \\
&= \sum_{t=2}^{T}\sum_{s<t}\sum_{r<t} \widetilde{\zeta}_s' C_{t-s,T}'(u)\, \widetilde{K}_{t,T}^2 C_{t-r,T}(u)\, \widetilde{\zeta}_r \\
&= \sum_{t=2}^{T}\sum_{s=1}^{t-1} \widetilde{\zeta}_s' C_{t-s,T}'(u)\, \widetilde{K}_{t,T}^2 C_{t-s,T}(u)\, \widetilde{\zeta}_s + \sum_{t=2}^{T}\sum_{s\neq r;s,r<t} \widetilde{\zeta}_s' C_{t-s,T}'(u)\, \widetilde{K}_{t,T}^2 C_{t-r,T}(u)\, \widetilde{\zeta}_r \equiv Z_1 + Z_2,
\end{aligned}
$$

where in the second equality $\widetilde{K}_{t,T}^2 = \frac{K_{h,tu}^2}{\frac{1}{T}\sum_{t=1}^{T} K_{h,tu}^2}$. Then it suffices to prove $Z_1 \xrightarrow{p} \sum_{a=1}^{p}\sum_{b=1}^{p} \eta_a\eta_b\Omega_{ab}$ and $Z_2 = o_p(1)$.

First we study Z_2. Note that $\mathbb{E}(Z_2) = 0$ and then following the proof of Lemma 2 in Lobato (1999),

$$
\begin{aligned}
&\mathbb{E}\left(Z_2^2\right) \\
&= \sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{s\neq r;s,r<t}\sum_{v\neq w;v,w<z} \mathbb{E}\left[\widetilde{\zeta}_s' C_{t-s,T}'(u)\, \widetilde{K}_{t,T}^2 C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{z-v,T}'(u)\, \widetilde{K}_{z,T}^2 C_{z-w,T}(u)\, \widetilde{\zeta}_w\right] \\
&= \sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{s\neq r;s,r<t}\sum_{v\neq w;v,w<z} \widetilde{K}_{t,T}^2 \widetilde{K}_{z,T}^2 \mathbb{E}\left[\operatorname{tr}\left(\widetilde{\zeta}_s' C_{t-s,T}'(u)\, C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{z-v,T}'(u)\, C_{z-w,T}(u)\, \widetilde{\zeta}_w\right)\right] \\
&= \sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{s\neq r;s,r<t}\sum_{v\neq w;v,w<z} \widetilde{K}_{t,T}^2 \widetilde{K}_{z,T}^2 \operatorname{vec}'\left(C_{t-s,T}'(u)\, C_{t-r,T}(u)\right) \mathbb{E}\left[\widetilde{\zeta}_r \widetilde{\zeta}_v' \otimes \widetilde{\zeta}_s \widetilde{\zeta}_w'\right] \operatorname{vec}\left(C_{z-v,T}'(u)\, C_{z-w,T}(u)\right),
\end{aligned}
$$

using the fact that $\operatorname{tr}(ABCD) = \operatorname{vec}(A')'(D' \otimes B)\operatorname{vec}(C)$. The expectation part above is given by

$$
\begin{aligned}
\mathbb{E}\left[\widetilde{\zeta}_r \widetilde{\zeta}_v' \otimes \widetilde{\zeta}_s \widetilde{\zeta}_w'\right] &= \frac{K_{h,ru}K_{h,vu}K_{h,su}K_{h,wu}}{\left(\frac{1}{T}\sum_{t=1}^{T} K_{h,tu}^2\right)^2} \mathbb{E}\left[\zeta_r\zeta_v' \otimes \zeta_s\zeta_w'\right] \\
&= \widetilde{K}_{r,T}\widetilde{K}_{v,T}\widetilde{K}_{s,T}\widetilde{K}_{w,T}\left[\left(I_{p+1} \otimes I_{p+1}\right)\mathbf{1}(r = v \neq s = w) + \widetilde{P}\mathbf{1}(r = w \neq s = v)\right]
\end{aligned}
$$

by Assumption 1(i), where $\widetilde{P}$ is defined as in Assumption 1(i) and (2.3.6) as a constant permutation matrix that has the same dimension as $I_{p+1} \otimes I_{p+1}$. Therefore we can decompose

$\mathbb{E}\left(Z_2^2\right)$ by

$$\mathbb{E}\left(Z_2^2\right) = \sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{\substack{s,r=1\\ s\neq r}}^{(z\wedge t)-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \mathrm{tr}\left(C_{t-s,T}'(u)\,C_{t-r,T}(u)\,C_{z-r,T}'(u)\,C_{z-s,T}(u)\right) \tag{A.2.19}$$

$$+\sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{\substack{s,r=1\\ s\neq r}}^{(z\wedge t)-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \mathrm{vec}\left(C_{t-s,T}'(u)\,C_{t-r,T}(u)\right)'\widetilde{P}\mathrm{vec}\left(C_{z-s,T}'(u)\,C_{z-r,T}(u)\right)$$

$$\equiv Z_{21}+Z_{22},$$

where $a\wedge b=\min(a,b)$. For Z_{21}, we have

$$\begin{aligned}
Z_{21} &= 2\sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{r>s}^{(z\wedge t)-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \mathrm{tr}\left(C_{t-s,T}'(u)\,C_{t-r,T}(u)\,C_{z-r,T}'(u)\,C_{z-s,T}(u)\right)\\
&= 2\sum_{t=2}^{T}\sum_{r>s}^{t-1} \widetilde{K}_{t,T}^4\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \mathrm{tr}\left(C_{t-s,T}'(u)\,C_{t-r,T}(u)\,C_{t-r,T}'(u)\,C_{t-s,T}(u)\right)\\
&+ 4\sum_{t=3}^{T}\sum_{z=2}^{t-1}\sum_{r>s}^{z-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \mathrm{tr}\left(C_{t-s,T}'(u)\,C_{t-r,T}(u)\,C_{z-r,T}'(u)\,C_{z-s,T}(u)\right)\\
&\equiv Z_{21,a}+Z_{21,b}.
\end{aligned}$$

It is easy to show that the order of $Z_{21,a}$ is not bigger than that of $Z_{21,b}$. So we can focus on $Z_{21,b}$ below. By Cauchy-Schwarz inequality,

$$Z_{21,b} \leq 4\sum_{t=3}^{T}\sum_{z=2}^{t-1}\sum_{r>s}^{z-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \left\|C_{t-s,T}'(u)\,C_{t-r,T}(u)\right\|\left\|C_{z-r,T}'(u)\,C_{z-s,T}(u)\right\| \tag{A.2.20}$$

$$\begin{aligned}
&\leq 4\sum_{t=3}^{T}\sum_{z=2}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\sum_{r=1}^{z-1}\widetilde{K}_{r,T}^2\left\|C_{z-r,T}(u)\right\|^2\sum_{r=1}^{z-1}\widetilde{K}_{r,T}^2\left\|C_{t-r,T}(u)\right\|^2\\
&\leq Qh^{-2}\sum_{t=3}^{T}\sum_{z=2}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\sum_{r=1}^{z-1}\left\|C_{z-r,T}(u)\right\|^2\sum_{r=1}^{z-1}\left\|C_{t-r,T}(u)\right\|^2\\
&= Qh^{-2}\sum_{t=3}^{T}\sum_{z=2}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\sum_{r=1}^{z-1}\left\|C_{r,T}(u)\right\|^2\sum_{r=t-z+1}^{t-1}\left\|C_{r,T}(u)\right\|^2\\
&\leq Qh^{-2}\left(\sum_{t=1}^{T}\left\|C_{t,T}(u)\right\|^2\right)\sum_{t=3}^{T}\widetilde{K}_{t,T}^2\sum_{z=2}^{t-1}\widetilde{K}_{z,T}^2\sum_{r=t-z+1}^{t-1}\left\|C_{r,T}(u)\right\|^2\\
&\leq QTh^{-3}\left(\sum_{t=1}^{T}\left\|C_{t,T}(u)\right\|^2\right)\left(\sum_{t=1}^{T}t\left\|C_{t,T}(u)\right\|^2\right),
\end{aligned}$$

where the third inequality holds for some absolute constant Q by Assumption 3(i) and

Riemann sum approximation, and the last inequality holds by the following:

$$\begin{aligned}
&\sum_{t=3}^{T}\sum_{z=2}^{t-1}\widetilde{K}_{t,T}^{2}\widetilde{K}_{z,T}^{2}\sum_{r=t-z+1}^{t-1}\left\|C_{r,T}\left(u\right)\right\|^{2}\\
&=\widetilde{K}_{2,T}^{2}\widetilde{K}_{3,T}^{2}\left\|C_{2,T}\left(u\right)\right\|^{2}\\
&+\widetilde{K}_{3,T}^{2}\widetilde{K}_{4,T}^{2}\left\|C_{2,T}\left(u\right)\right\|^{2}+\left(\widetilde{K}_{2,T}^{2}\widetilde{K}_{4,T}^{2}+\widetilde{K}_{3,T}^{2}\widetilde{K}_{4,T}^{2}\right)\left\|C_{3,T}\left(u\right)\right\|^{2}\\
&+\widetilde{K}_{4,T}^{2}\widetilde{K}_{5,T}^{2}\left\|C_{2,T}\left(u\right)\right\|^{2}+\left(\widetilde{K}_{3,T}^{2}\widetilde{K}_{5,T}^{2}+\widetilde{K}_{4,T}^{2}\widetilde{K}_{5,T}^{2}\right)\left\|C_{3,T}\left(u\right)\right\|^{2}\\
&+\left(\widetilde{K}_{2,T}^{2}\widetilde{K}_{5,T}^{2}+\widetilde{K}_{3,T}^{2}\widetilde{K}_{5,T}^{2}+\widetilde{K}_{4,T}^{2}\widetilde{K}_{5,T}^{2}\right)\left\|C_{4,T}\left(u\right)\right\|^{2}\\
&\dots\\
&+\widetilde{K}_{T-1,T}^{2}\widetilde{K}_{T,T}^{2}\left\|C_{2,T}\left(u\right)\right\|^{2}+\cdots+\left(\widetilde{K}_{2,T}^{2}\widetilde{K}_{T,T}^{2}+\cdots+\widetilde{K}_{T-1,T}^{2}\widetilde{K}_{T,T}^{2}\right)\left\|C_{T-1,T}\left(u\right)\right\|^{2}\\
&\leq C\sum_{t=3}^{T}\left\|C_{t-1,T}\left(u\right)\right\|^{2}(t-2)\sum_{s=2}^{T-t+2}\widetilde{K}_{s,T}^{4}\\
&\leq C\left(\sum_{t=1}^{T}\widetilde{K}_{t,T}^{4}\right)\left(\sum_{t=1}^{T}t\left\|C_{t,T}\left(u\right)\right\|^{2}\right)\leq CTh^{-1}\sum_{t=1}^{\lfloor T/2\rfloor}t\left\|C_{t,T}\left(u\right)\right\|^{2},
\end{aligned}$$

for some constant C, where in the last inequality $\lfloor X\rfloor$ is defined as the largest integer that is smaller than X. Note that our $C_{t,T}\left(u\right)$ is numerically identical to that in the proof of Theorem 1 in Nielsen (2005) except the dependence on u and an extra multiplier $\sqrt{h}$, and this dependence makes the asymptotic order of $C_{t,T}\left(u\right)$ analogous to the corresponding order in Nielsen (2005, pp. 301-302) as

$$\left\|C_{t,T}\left(u\right)\right\|=O\left(\frac{\sqrt{Mh}}{T}+\frac{\sqrt{h}}{t\sqrt{M}}\right) \tag{A.2.21}$$

and

$$\sum_{t=1}^{T}\left\|C_{t,T}\left(u\right)\right\|^{2}=O\left(\sum_{t=1}^{\lfloor T/M\rfloor}\frac{Mh}{T^{2}}+\sum_{t=\lfloor T/M\rfloor+1}^{T}\frac{h}{t^{2}M}\right)=O\left(\frac{h}{T}\right),$$

using the cut-off $t=\lfloor T/M\rfloor$ as in Nielsen (2005). Then for $h^{-3}\sum_{t=1}^{\lfloor T/2\rfloor}t\left\|C_{t,T}\left(u\right)\right\|^{2}$ we use a different cut-off $t=\lfloor T/M\rfloor$ from Nielsen (2005) and the sum is given by

$$\begin{aligned}
h^{-3}\sum_{t=1}^{\lfloor T/2\rfloor}t\left\|C_{t,T}\left(u\right)\right\|^{2}&=O\left(h^{-2}\left(\sum_{t=1}^{\lfloor T/M\rfloor}\frac{tM}{T^{2}}+\sum_{t=\lfloor T/M\rfloor+1}^{\lfloor T/2\rfloor}\frac{1}{tM}\right)\right)\\
&=O\left(\frac{M}{T^{2}h^{2}}\left(T/M\right)^{2}+\frac{1}{Mh^{2}}\left(\log T-\log\frac{T}{M}\right)\right)\\
&=O\left(\frac{\log M}{Mh^{2}}\right)=o\left(1\right)
\end{aligned}$$

by Assumption 4(iii). Note that $\lfloor T/M \rfloor$ is an appropriate cut-off because

$$\frac{T}{2} > \frac{T}{M} > 1$$

in asymptotics. Therefore we prove the negligibility for Z_{21}. Next for Z_{22} we have

$$\begin{aligned}
&|Z_{22}| \\
&\leq \sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{s\neq r;s,r<t}^{(z\wedge t)-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \left\|\mathrm{vec}\left(C_{t-s,T}'(u)\,C_{t-r,T}(u)\right)\right\| \left\|\widetilde{P}\right\| \left\|\mathrm{vec}\left(C_{z-s,T}'(u)\,C_{z-r,T}(u)\right)\right\| \\
&= \left\|\widetilde{P}\right\| \sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{s\neq r;s,r<t}^{(z\wedge t)-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \left\|\mathrm{vec}\left(C_{t-s,T}'(u)\,C_{t-r,T}(u)\right)\right\| \left\|\mathrm{vec}\left(C_{z-s,T}'(u)\,C_{z-r,T}(u)\right)\right\| \\
&\leq 2\left\|\widetilde{P}\right\| \sum_{t=2}^{T}\sum_{z=2}^{T}\sum_{r>s}^{(z\wedge t)-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \left\|C_{t-s,T}'(u)\,C_{t-r,T}(u)\right\| \left\|C_{z-s,T}'(u)\,C_{z-r,T}(u)\right\| \\
&= 2\left\|\widetilde{P}\right\| \sum_{t=2}^{T}\sum_{r>s}^{t-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \left\|C_{t-s,T}'(u)\,C_{t-r,T}(u)\right\|^2 \\
&+ 4\left\|\widetilde{P}\right\| \sum_{t=3}^{T}\sum_{z=2}^{t-1}\sum_{r>s}^{z-1} \widetilde{K}_{t,T}^2\widetilde{K}_{z,T}^2\widetilde{K}_{r,T}^2\widetilde{K}_{s,T}^2 \left\|C_{t-s,T}'(u)\,C_{t-r,T}(u)\right\| \left\|C_{z-s,T}'(u)\,C_{z-r,T}(u)\right\| \\
&\equiv Z_{22,a} + Z_{22,b},
\end{aligned}$$

where the third inequality holds by the fact that $\|\mathrm{vec}(X)\| = \|X\|$ for any matrix X. Note that $\left\|\widetilde{P}\right\|$ is finite as a permutation matrix. Therefore as we just analyzed, $Z_{22,a}$ is smaller than $Z_{22,b}$, and $Z_{22,b}$ has the same bound as in (A.2.20), which implies that Z_{22} is negligible as well.

Next, we prove $Z_1 = \sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{\zeta}_s' C_{t-s,T}'(u)\,\widetilde{K}_{t,T}^2 C_{t-s,T}(u)\,\widetilde{\zeta}_s \xrightarrow{p} \sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\Omega_{ab}$. For $\mathbb{E}(Z_1)$, we have

$$\begin{aligned}
\mathbb{E}(Z_1) &= \sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2 \mathrm{tr}\left[C_{t-s,T}'(u)\,C_{t-s,T}(u)\,\mathbb{E}\left(\widetilde{\zeta}_s\widetilde{\zeta}_s'\right)\right] \\
&= \sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2 \mathrm{tr}\left[C_{t-s,T}'(u)\,C_{t-s,T}(u)\right] \\
&= \sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{h\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2}{4\pi^2T^2M}\sum_{j,k=1}^{M}\mathrm{tr}\left[\theta_j'\theta_k\right]\cos\left((t-s)\,\lambda_j\right)\cos\left((t-s)\,\lambda_k\right) \\
&= \sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{h\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2}{4\pi^2T^2M}\sum_{j=1}^{M}\mathrm{tr}\left[\theta_j'\theta_j\right]\cos^2\left((t-s)\,\lambda_j\right) \\
&+ \sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{h\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2}{4\pi^2T^2M}\sum_{j\neq k}\mathrm{tr}\left[\theta_j'\theta_k\right]\cos\left((t-s)\,\lambda_j\right)\cos\left((t-s)\,\lambda_k\right) \\
&\equiv EZ_{11} + EZ_{12}.
\end{aligned}$$

For both EZ_{11} and EZ_{12}, we can firstly consider the summation over time domain, under either $j = k$ or $j \neq k$, which is given by

$$\begin{aligned}
&\frac{h}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\cos\left((t-s)\lambda_j\right)\cos\left((t-s)\lambda_k\right) \qquad \text{(A.2.22)}\\
&=\frac{h}{4T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\left(e^{i(t-s)\lambda_j}+e^{-i(t-s)\lambda_j}\right)\left(e^{i(t-s)\lambda_k}+e^{-i(t-s)\lambda_k}\right)\\
&=\frac{h}{4T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\left(e^{i(t-s)\lambda_{j+k}}+e^{-i(t-s)\lambda_{j-k}}+e^{i(t-s)\lambda_{j-k}}+e^{-i(t-s)\lambda_{j+k}}\right)\\
&=\frac{h}{8T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\left(e^{i(t-s)\lambda_{j+k}}+e^{-i(t-s)\lambda_{j-k}}+e^{i(t-s)\lambda_{j-k}}+e^{-i(t-s)\lambda_{j+k}}\right)-\frac{h}{8T^2}\sum_{t=1}^{T}\widetilde{K}_{t,T}^4\\
&=\frac{h}{4T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\left(e^{i(t-s)\lambda_{j+k}}+e^{i(t-s)\lambda_{j-k}}\right)-\frac{h}{8T^2}\sum_{t=1}^{T}\widetilde{K}_{t,T}^4\\
&=\frac{h}{4T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\left(e^{i(t-s)\lambda_{j+k}}+e^{i(t-s)\lambda_{j-k}}\right)+O\left(\frac{1}{T}\right)
\end{aligned}$$

Note that $\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2 e^{i(t-s)\lambda_{j+k}}$ is the periodogram of $\widetilde{K}_{t,T}^2$ at frequency λ_{j+k}, and it is thus equal to $\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2 e^{-i(t-s)\lambda_{j+k}}$, and $\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2 e^{i(t-s)\lambda_{j-k}}$ has the similar argument. Then we can focus on the DFT of $\widetilde{K}_{t,T}^2$ at frequencies λ_{j+k} and λ_{j-k}, in which we consider the latter for instance. Note that when $j = k$, it is easy to see that $\frac{\sqrt{h}}{T}\sum_{t=1}^{T}\widetilde{K}_{t,T}^2 e^{it\lambda_{j-k}} = \frac{\sqrt{h}}{T}\sum_{t=1}^{T}\widetilde{K}_{t,T}^2 = \sqrt{h}$; and when $j \neq k$ we denote $q = j - k$ and it follows by Assumption 3(iv) that

$$\begin{aligned}
\left|\frac{\sqrt{h}}{T}\sum_{t=1}^{T}\widetilde{K}_{t,T}^2 e^{it\lambda_q}\right| &= \frac{\sqrt{h}}{\sum_{t=1}^{T}k^2\left(\frac{t-Tu}{Th}\right)}\left|\sum_{t=1}^{T}k^2\left(\frac{t-Tu}{Th}\right)e^{it\lambda_q}\right|\\
&=\frac{\sqrt{h}}{Th\kappa_{02}+o(1)}\left|\sum_{t=1}^{T}k^2\left(\frac{t-Tu}{Th}\right)e^{it\lambda_q}\right| \to \frac{\theta(u,q)}{\kappa_{02}} < \infty \qquad \text{(A.2.23)}
\end{aligned}$$

as $T \to \infty$. And the same reasoning can hold for DFT of $\widetilde{K}_{t,T}^2$ at frequencies λ_{j+k} as well. Then since a periodogram is the squared modulus of a DFT, we can see that

$$\frac{h}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2 e^{i(t-s)\lambda_{j-k}} \to \frac{\theta(u,j-k)}{\kappa_{02}}.$$

And a similar consequence holds for $\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2 e^{i(t-s)\lambda_{j+k}}$, which then altogether implies that

$$\begin{aligned}
&\frac{h}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\cos\left((t-s)\lambda_j\right)\cos\left((t-s)\lambda_k\right)\\
&=\frac{h}{4T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\left(e^{i(t-s)\lambda_{j+k}}+e^{i(t-s)\lambda_{j-k}}\right)+O\left(\frac{1}{T}\right)
\end{aligned}$$

$$= \frac{\theta(u, j-k) + \theta(u, j-k)}{4\kappa_{02}} \mathbf{1}(j \neq k) + o\left(\frac{1}{M}\right) + O\left(\frac{1}{T}\right)$$
$$\equiv \left(\frac{\theta^*(u, j, k)}{4\kappa_{02}} + o\left(\frac{1}{M}\right)\right) \mathbf{1}(j \neq k).$$

Then it is sufficient to consider the asymptotics of EZ_{12} with summation over j and k, which is given by

$$EZ_{12} = \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{h\widetilde{K}_{t,T}^2 \widetilde{K}_{s,T}^2}{4\pi^2 T^2 M} \sum_{j \neq k} \mathrm{tr}\left[\theta_j' \theta_k\right] \cos\left((t-s)\lambda_j\right) \cos\left((t-s)\lambda_k\right) \tag{A.2.24}$$
$$= \frac{\lambda_M^{d_{X_a}(u)+d_{X_b}(u)+2d_\varepsilon(u)-4\delta(u)}}{4\pi^2 \kappa_{02} M} \sum_{a=1}^{p} \sum_{b=1}^{p} \eta_a \eta_b \sum_{j \neq k; j+k \leq \overline{\theta}(u); j-k \leq \overline{\theta}(u)} \theta^*(u, j, k)$$
$$\cdot \lambda_j^{2\delta(u)} \lambda_k^{2\delta(u)} \mathrm{tr}\left[\mathrm{Re}\left(A_{X_a,j}' \overline{A}_{\varepsilon,j} + A_{\varepsilon,j}' \overline{A}_{X_a,j}\right) \mathrm{Re}\left(A_{X_b,k}' \overline{A}_{\varepsilon,k} + A_{\varepsilon,k}' \overline{A}_{X_b,k}\right)\right]$$
$$+ \frac{\lambda_M^{2d_{X_a}(u)+2d_\varepsilon(u)-4\delta(u)}}{4\pi^2 \kappa_{02} M} \sum_{a=1}^{p} \sum_{b=1}^{p} \eta_a \eta_b \sum_{j \neq k} o\left(\frac{1}{M}\right)$$
$$\cdot \lambda_j^{2\delta(u)} \lambda_k^{2\delta(u)} \mathrm{tr}\left[\mathrm{Re}\left(A_{X_a,j}' \overline{A}_{\varepsilon,j} + A_{\varepsilon,j}' \overline{A}_{X_a,j}\right) \mathrm{Re}\left(A_{X_b,k}' \overline{A}_{\varepsilon,k} + A_{\varepsilon,k}' \overline{A}_{X_b,k}\right)\right]$$

where in the second equality we take into consideration the definition of the limit $\theta(u, j)$, note that the finiteness of $\overline{\theta}(u)$ guarantees the validity of order $\frac{1}{M}$. And within the trace operator above, we have $\mathrm{Re}\left(A_{X_a,j}' \overline{A}_{\varepsilon,j} + A_{\varepsilon,j}' \overline{A}_{X_a,j}\right)$ given by

$$\mathrm{Re}\left(A_{X_a,j}' \overline{A}_{\varepsilon,j} + A_{\varepsilon,j}' \overline{A}_{X_a,j}\right) = \mathrm{Re}\left(A_{X_a,j}^* A_{\varepsilon,j} + A_{\varepsilon,j}^* A_{X_a,j}\right)$$
$$= \mathrm{Re}\left[\begin{pmatrix} 0_{p\times p} & \widetilde{A}_{X_a,j}^* \widetilde{A}_{\varepsilon,j} \\ 0_{1\times p} & 0_{1\times 1} \end{pmatrix} + \begin{pmatrix} 0_{p\times p} & 0_{p\times 1} \\ \widetilde{A}_{X_a,j} \widetilde{A}_{\varepsilon,j}^* & 0_{1\times 1} \end{pmatrix}\right]$$
$$= \begin{pmatrix} 0_{p\times p} & \mathrm{Re}\left(\widetilde{A}_{X_a,j}^* \widetilde{A}_{\varepsilon,j}\right) \\ \mathrm{Re}\left(\widetilde{A}_{X_a,j} \widetilde{A}_{\varepsilon,j}^*\right) & 0_{1\times 1} \end{pmatrix}$$

from Assumption 1(i) and the fact that $\mathrm{Re}(X) = \mathrm{Re}\left(\overline{X}\right)$ for any complex vector X with conjugate $\overline{X}$, and thus the similar conclusion holds for $\mathrm{Re}\left(A_{X_a,k}' \overline{A}_{\varepsilon,k} + A_{\varepsilon,k}' \overline{A}_{X_a,k}\right)$. Therefore

$$\mathrm{tr}\left[\mathrm{Re}\left(A_{X_a,j}' \overline{A}_{\varepsilon,j} + A_{\varepsilon,j}' \overline{A}_{X_a,j}\right) \mathrm{Re}\left(A_{X_b,k}' \overline{A}_{\varepsilon,k} + A_{\varepsilon,k}' \overline{A}_{X_b,k}\right)\right]$$
$$= \mathrm{tr}\left[\mathrm{Re}\left(A_{X_a,j}^* A_{\varepsilon,j} + A_{\varepsilon,j}^* A_{X_a,j}\right) \mathrm{Re}\left(A_{X_b,k}^* A_{\varepsilon,k} + A_{\varepsilon,k}^* A_{X_b,k}\right)\right]$$
$$= \mathrm{tr}\left[\begin{pmatrix} 0_{p\times p} & \mathrm{Re}\left(\widetilde{A}_{X_a,j}^* \widetilde{A}_{\varepsilon,j}\right) \\ \mathrm{Re}\left(\widetilde{A}_{X_a,j} \widetilde{A}_{\varepsilon,j}^*\right) & 0_{1\times 1} \end{pmatrix} \begin{pmatrix} 0_{p\times p} & \mathrm{Re}\left(\widetilde{A}_{X_b,k}^* \widetilde{A}_{\varepsilon,k}\right) \\ \mathrm{Re}\left(\widetilde{A}_{X_b,k} \widetilde{A}_{\varepsilon,k}^*\right) & 0_{1\times 1} \end{pmatrix}\right]$$
$$= \mathrm{tr}\left[\begin{pmatrix} \mathrm{Re}\left(\widetilde{A}_{X_a,j}^* \widetilde{A}_{\varepsilon,j}\right) \mathrm{Re}\left(\widetilde{A}_{X_b,k} \widetilde{A}_{\varepsilon,k}^*\right) & 0_{p\times 1} \\ 0_{1\times p} & \mathrm{Re}\left(\widetilde{A}_{X_a,j} \widetilde{A}_{\varepsilon,j}^*\right) \mathrm{Re}\left(\widetilde{A}_{X_b,k}^* \widetilde{A}_{\varepsilon,k}\right) \end{pmatrix}\right]$$

$$= \mathrm{tr}\left[\mathrm{Re}\left(\widetilde{A}^*_{X_a,j}\widetilde{A}_{\varepsilon,j}\right)\mathrm{Re}\left(\widetilde{A}_{X_b,k}\widetilde{A}^*_{\varepsilon,k}\right)\right] + \mathrm{Re}\left(\widetilde{A}_{X_a,j}\widetilde{A}^*_{\varepsilon,j}\right)\mathrm{Re}\left(\widetilde{A}^*_{X_b,k}\widetilde{A}_{\varepsilon,k}\right)$$
$$= \mathrm{Re}\left(\widetilde{A}_{X_b,k}\widetilde{A}^*_{\varepsilon,k}\right)\mathrm{Re}\left(\widetilde{A}^*_{X_a,j}\widetilde{A}_{\varepsilon,j}\right) + \mathrm{Re}\left(\widetilde{A}_{X_a,j}\widetilde{A}^*_{\varepsilon,j}\right)\mathrm{Re}\left(\widetilde{A}^*_{X_b,k}\widetilde{A}_{\varepsilon,k}\right).$$

Denote the l-th element of $\widetilde{A}_{X_a,k}$ as $\widetilde{A}_{X_a,k(l)}$, then we have

$$\mathrm{Re}\left(\widetilde{A}_{X_b,k}\widetilde{A}^*_{\varepsilon,k}\right)\mathrm{Re}\left(\widetilde{A}^*_{X_a,j}\widetilde{A}_{\varepsilon,j}\right) = \sum_{l=1}^{p}\mathrm{Re}\left(\widetilde{A}_{X_b,k(l)}\widetilde{A}^*_{\varepsilon,k}\right)\mathrm{Re}\left(\widetilde{A}^*_{X_a,j(l)}\widetilde{A}_{\varepsilon,j}\right).$$

Note that by Assumption 1(i) and 2(i), $f_\varepsilon(u,\lambda_j) = \frac{1}{2\pi}\left|\widetilde{A}_{\varepsilon,j}\right|^2$ and $f_{X_a}(u,\lambda_j) = \frac{1}{2\pi}\left|\widetilde{A}_{X_a,j}\right|^2$. Then by Lobato (1997, pp. 151), there exist a nonempty subset $L \subseteq \{1,\dots,p\}$ such that $\frac{1}{2\pi}\left|\widetilde{A}_{X_a,j(l)}\right|^2 \sim C_l\lambda_j^{-2d_{X_a}(u)}$ for all $l \in L$ and $\sum_{l\in L} C_l = G_{X_{aa}}$, and for the rest of the arguments $m \in \{1,\dots,p\}\backslash L$, $\frac{1}{2\pi}\left|\widetilde{A}_{X_a,j(m)}\right|^2 = o\left(\lambda_j^{-2d_{X_a}(u)}\right)$. Then take $\mathrm{Re}\left(\widetilde{A}^*_{X_a,j}\widetilde{A}_{\varepsilon,j}\right)$ for instance, we can see that

$$\mathrm{Re}\left(\widetilde{A}^*_{X_a,j}\widetilde{A}_{\varepsilon,j}\right) = \mathrm{Re}\left(\widetilde{A}_{X_a,j}\right)\mathrm{Re}\left(\widetilde{A}_{\varepsilon,j}\right) - \mathrm{Im}\left(\widetilde{A}_{X_a,j}\right)\mathrm{Im}\left(\widetilde{A}_{\varepsilon,j}\right).$$

And note that by Assumption 2(ii) $\mathrm{Im}\left(\widetilde{A}_{X_a,j}\right) = O\left(\lambda_j^{\gamma-2d_{X_a}(u)}\right)$ and $\mathrm{Im}\left(\widetilde{A}_{\varepsilon,j}\right) = O\left(\lambda_j^{\gamma-2d_\varepsilon(u)}\right)$, which implies that the $\mathrm{Re}\left(\widetilde{A}_{X_a,j}\right)$ and $\mathrm{Re}\left(\widetilde{A}_{\varepsilon,j}\right)$ are the leading terms in $\left|\widetilde{A}_{X_a,j}\right|^2$ and $\left|\widetilde{A}_{\varepsilon,j}\right|^2$ respectively, therefore

$$\begin{aligned}\mathrm{Re}\left(\widetilde{A}^*_{X_a,j}\widetilde{A}_{\varepsilon,j}\right) &= \left|\widetilde{A}_{X_a,j}\right|\left|\widetilde{A}_{\varepsilon,j}\right| + O\left(\lambda_j^{2\gamma-2d_{X_a}(u)-2d_\varepsilon(u)}\right)\\ &\sim 2\pi G^{\frac{1}{2}}_{X_{aa}}G^{\frac{1}{2}}_{\varepsilon}\lambda_j^{-d_{X_a}(u)-d_\varepsilon(u)}.\end{aligned}$$

By the same reasoning, $\mathrm{Re}\left(\widetilde{A}_{X_b,k}\widetilde{A}^*_{\varepsilon,k}\right) \sim 2\pi G^{\frac{1}{2}}_{X_{bb}}G^{\frac{1}{2}}_{\varepsilon}\lambda_k^{-d_{X_b}(u)-d_\varepsilon(u)}$, which implies that

$$\begin{aligned}&\mathrm{tr}\left[\mathrm{Re}\left(A'_{X_a,j}\overline{A}_{\varepsilon,j} + A'_{\varepsilon,j}\overline{A}_{X_a,j}\right)\mathrm{Re}\left(A'_{X_b,k}\overline{A}_{\varepsilon,k} + A'_{\varepsilon,k}\overline{A}_{X_b,k}\right)\right]\\ &\sim 8\pi^2 G^{\frac{1}{2}}_{X_{aa}}G^{\frac{1}{2}}_{X_{bb}}G_\varepsilon\lambda_j^{-d_{X_a}(u)-d_\varepsilon(u)}\lambda_k^{-d_{X_b}(u)-d_\varepsilon(u)}.\end{aligned}$$

Substituting this result into (A.2.24) we have

$$\begin{aligned}EZ_{12} =& \sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{h\widetilde{K}^2_{t,T}\widetilde{K}^2_{s,T}}{4\pi^2T^2M}\sum_{j,k=1}^{M}\mathrm{tr}\left[\theta'_j\theta_k\right]\cos\left((t-s)\lambda_j\right)\cos\left((t-s)\lambda_k\right)\\ &\sim \frac{2G^{\frac{1}{2}}_{X_{aa}}G^{\frac{1}{2}}_{X_{bb}}G_\varepsilon\lambda_M^{d_{X_a}(u)+d_{X_b}(u)+2d_\varepsilon(u)-4\delta(u)}}{\kappa_{02}M}\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\\ &\cdot\sum_{j\neq k;j+k\leq\overline{\theta}(u);j-k\leq\overline{\theta}(u)}\theta^*(u,j,k)\,\lambda_j^{2\delta(u)-d_{X_a}(u)-d_\varepsilon(u)}\lambda_k^{2\delta(u)-d_{X_b}(u)-d_\varepsilon(u)}\end{aligned}$$

$$\to \sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\Theta^*\left(\frac{G_{X_{aa}}^{\frac{1}{2}}G_{X_{bb}}^{\frac{1}{2}}G_{\varepsilon}}{\left(1-d_{X_a}(u)-d_{\varepsilon}(u)+2\delta(u)\right)}\right).$$

Then it remains to analyze the order of EZ_{11}, which can be done in the same manner as above. Note that in EZ_{11},

$$\cos^2\left(\left(t-s\right)\lambda_j\right)=\frac{1}{2}\left[\cos\left(2\left(t-s\right)\lambda_j\right)+1\right].$$

Then consider the summation over time domain first, we have

$$\begin{aligned}
&\frac{h}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\cos^2\left(\left(t-s\right)\lambda_j\right)\\
&=\frac{h}{2T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\left(\cos\left(2\left(t-s\right)\lambda_j\right)+1\right)\\
&=\frac{h}{2T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\cos\left(2\left(t-s\right)\lambda_j\right)+\frac{h}{2T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\equiv EZ_{11,a}+EZ_{11,b},
\end{aligned}$$

where firstly $EZ_{11,b}$ is given by

$$\begin{aligned}
EZ_{11,b}&=\frac{h}{2T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2=\frac{h}{2T^2}\left[\left(\sum_{t=1}^{T}\widetilde{K}_{t,T}^2\right)^2-\sum_{t=1}^{T}\widetilde{K}_{t,T}^4\right]\\
&=\frac{h}{2T^2}\left(\sum_{t=1}^{T}\frac{K_{h,tu}^2}{\frac{1}{T}\sum_{t=1}^{T}K_{h,tu}^2}\right)^2-\frac{h}{2T^2}\sum_{t=1}^{T}\frac{K_{h,tu}^4}{\left(\frac{1}{T}\sum_{t=1}^{T}K_{h,tu}^2\right)^2}\\
&=O\left(h+T^{-1}\right),
\end{aligned}$$

and then $EZ_{11,a}$ is given by

$$\begin{aligned}
EZ_{11,a}&=\frac{h}{2T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\cos\left(2\left(t-s\right)\lambda_j\right)\\
&=\frac{h}{4T^2}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t,T}^2\widetilde{K}_{s,T}^2\cos\left(2\left(t-s\right)\lambda_j\right)-\sum_{t=1}^{T}\widetilde{K}_{t,T}^4\right]\\
&=\frac{h}{4T^2}\operatorname{Re}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{K_{h,tu}^2K_{h,su}^2}{\left(\frac{1}{T}\sum_{t=1}^{T}K_{h,tu}^2\right)^2}e^{i(t-s)2\lambda_j}\right)+O\left(T^{-1}\right)\\
&\le\theta\left(u,2j\right)+o\left(\frac{1}{M}\right),
\end{aligned}$$

by Assumption 3(iv). Thus $EZ_{11,a}+EZ_{11,b}\le\theta\left(u,2j\right)+O\left(h\right)$, and this order holds uniformly over j. Then given the properties of function $\theta(u,j)$, we have as $T\to\infty$,

$$
\begin{aligned}
EZ_{11} &= \sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{h\widetilde{K}_{t,T}^{2}\widetilde{K}_{s,T}^{2}}{4\pi^{2}T^{2}M}\sum_{j=1}^{M}\operatorname{tr}\left[\theta_j^{'}\theta_j\right]\cos^{2}\left((t-s)\lambda_j\right)\\
&\le \frac{2G_{X_{ab}}G_{\varepsilon}\lambda_M^{d_{X_a}(u)+d_{X_b}(u)+2d_{\varepsilon}(u)-4\delta(u)}}{\kappa_{02}M}\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\sum_{j\le\overline{\theta}(u)}\theta(u,2j)\lambda_j^{4\delta(u)-d_{X_a}(u)-d_{X_b}(u)-2d_{\varepsilon}(u)}\\
&+O(h)\frac{2G_{X_{ab}}G_{\varepsilon}\lambda_M^{d_{X_a}(u)+d_{X_b}(u)+2d_{\varepsilon}(u)-4\delta(u)}}{\kappa_{02}M}\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\sum_{j=1}^{M}\lambda_j^{4\delta(u)-d_{X_a}(u)-d_{X_b}(u)-2d_{\varepsilon}(u)}\\
&= \frac{2G_{X_{ab}}G_{\varepsilon}}{\kappa_{02}M}\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\sum_{j\le\overline{\theta}(u)}\theta(u,2j)\,\frac{j}{M}^{4\delta(u)-d_{X_a}(u)-d_{X_b}(u)-2d_{\varepsilon}(u)}+O(h)=o(1)
\end{aligned}
$$

using the conclusion in Nielsen (2005, pp. 294) and the fact that $\overline{\theta}(u)$ is finite. Therefore combining what we have derived for EZ_{11} and EZ_{12}, we can conclude that

$$
\mathbb{E}(Z_1)\to\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\Theta^{*}\left(\frac{G_{X_{aa}}^{\frac{1}{2}}G_{X_{bb}}^{\frac{1}{2}}G_{\varepsilon}}{\left(1-d_{X_a}(u)-d_{\varepsilon}(u)+2\delta(u)\right)}\right)=\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\Omega_{ab}.
$$

Next to prove the convergence in probability we study $\mathbb{E}\left(Z_1^2\right)$. Note that

$$
\begin{aligned}
&\mathbb{E}\left(Z_1^2\right) \qquad \text{(A.2.25)}\\
&=\mathbb{E}\left|\sum_{t=2}^{T}\sum_{s=1}^{t-1}\widetilde{\zeta}_s^{\prime}C_{t-s,T}^{\prime}(u)\widetilde{K}_{t,T}^{2}C_{t-s,T}(u)\widetilde{\zeta}_s\right|^{2}\\
&=\mathbb{E}\left(\sum_{t_1,t_2=2}^{T}\sum_{s_1=1}^{t_1-1}\sum_{s_2=1}^{t_2-1}\widetilde{\zeta}_{s_1}^{\prime}C_{t_1-s_1,T}^{\prime}(u)\widetilde{K}_{t_1,T}^{2}C_{t_1-s_1,T}(u)\widetilde{\zeta}_{s_1}\widetilde{\zeta}_{s_2}^{\prime}C_{t_2-s_2,T}^{\prime}(u)\widetilde{K}_{t_2,T}^{2}C_{t_2-s_2,T}(u)\widetilde{\zeta}_{s_2}\right)\\
&=\sum_{t_1,t_2=2}^{T}\sum_{s_1=1}^{t_1-1}\sum_{s_2=1}^{t_2-1}\widetilde{K}_{t_1,T}^{2}\widetilde{K}_{t_2,T}^{2}\mathbb{E}\left(\operatorname{tr}\left[C_{t_1-s_1,T}^{\prime}(u)C_{t_1-s_1,T}(u)\widetilde{\zeta}_{s_1}\widetilde{\zeta}_{s_2}^{\prime}C_{t_2-s_2,T}^{\prime}(u)C_{t_2-s_2,T}(u)\widetilde{\zeta}_{s_2}\widetilde{\zeta}_{s_1}^{\prime}\right]\right)\\
&=\sum_{t_1,t_2=2}^{T}\sum_{s_1=1}^{t_1-1}\sum_{s_2=1}^{t_2-1}\widetilde{K}_{t_1,T}^{2}\widetilde{K}_{t_2,T}^{2}\operatorname{vec}\left(C_{t_2-s_2,T}^{\prime}(u)C_{t_2-s_2,T}(u)\right)^{\prime}\mathbb{E}\left(\widetilde{\zeta}_{s_2}\widetilde{\zeta}_{s_1}^{\prime}\otimes\widetilde{\zeta}_{s_2}\widetilde{\zeta}_{s_1}^{\prime}\right)\operatorname{vec}\left(C_{t_1-s_1,T}^{\prime}(u)C_{t_1-s_1,T}(u)\right)\\
&=\sum_{t_1,t_2=2}^{T}\sum_{s_1=1}^{t_1-1}\sum_{s_2=1}^{t_2-1}\widetilde{K}_{t_1,T}^{2}\widetilde{K}_{t_2,T}^{2}\widetilde{K}_{s_1,T}^{2}\widetilde{K}_{s_2,T}^{2}\\
&\times\operatorname{vec}\left(C_{t_2-s_2,T}^{\prime}(u)C_{t_2-s_2,T}(u)\right)^{\prime}\mathbb{E}\left(\zeta_{s_2}\zeta_{s_1}^{\prime}\otimes\zeta_{s_2}\zeta_{s_1}^{\prime}\right)\operatorname{vec}\left(C_{t_1-s_1,T}^{\prime}(u)C_{t_1-s_1,T}(u)\right)\\
&=\sum_{t_1,t_2=2}^{T}\sum_{s=1}^{(t_1-1)\wedge(t_2-1)}\widetilde{K}_{t_1,T}^{2}\widetilde{K}_{t_2,T}^{2}\widetilde{K}_{s,T}^{4}\\
&\times\operatorname{vec}\left(C_{t_2-s_2,T}^{\prime}(u)C_{t_2-s_2,T}(u)\right)^{\prime}\mathbb{E}\left(\zeta_s\zeta_s^{\prime}\otimes\zeta_s\zeta_s^{\prime}\right)\operatorname{vec}\left(C_{t_1-s_1,T}^{\prime}(u)C_{t_1-s_1,T}(u)\right)\\
&+\sum_{t_1,t_2=2}^{T}\sum_{\substack{s_1,s_2=1\\ s_1\neq s_2}}^{(t_1-1)\wedge(t_2-1)}\widetilde{K}_{t_1,T}^{2}\widetilde{K}_{t_2,T}^{2}\widetilde{K}_{s_1,T}^{2}\widetilde{K}_{s_2,T}^{2}\\
&\times\left(\operatorname{vec}\left(C_{t_2-s_2,T}^{'}(u)C_{t_2-s_2,T}(u)\right)^{'}\mathbb{E}\left(\zeta_{s_2}\zeta_{s_1}^{'}\otimes\zeta_{s_2}\zeta_{s_1}^{'}\right)\operatorname{vec}\left(C_{t_1-s_1,T}^{'}(u)C_{t_1-s_1,T}(u)\right)\right)
\end{aligned}
$$

$$\equiv SZ_1 + SZ_2.$$

Firstly for SZ_2, we have $\mathbb{E}\left(\zeta_{s_2}\zeta_{s_1}^{'} \otimes \zeta_{s_2}\zeta_{s_1}^{'}\right) = \widetilde{C}$ by Assumption 1(i) with the sparse constant matrix $\widetilde{C}$ is given in detail by

$$\widetilde{C} = \begin{pmatrix} \boldsymbol{e}_{11} & \boldsymbol{e}_{12} & \cdots & \boldsymbol{e}_{1,p+1} \\ \vdots & \ddots & & \vdots \\ \boldsymbol{e}_{p+1,1} & \boldsymbol{e}_{p+1,2} & \cdots & \boldsymbol{e}_{p+1,p+1} \end{pmatrix},$$

where $\boldsymbol{e}_{ij}$ is a $(p+1) \times (p+1)$ matrix with all elements equal to zero except the (i,j)-th one. Therefore

$$\begin{aligned}
&SZ_2 \\
&= \sum_{t_1=2}^{T}\sum_{t_2=2}^{T}\sum_{\substack{s_1,s_2=1\\ s_1\neq s_2}}^{(t_1-1)\wedge(t_2-1)} \widetilde{K}_{t_1,T}^2\widetilde{K}_{t_2,T}^2\widetilde{K}_{s_1,T}^2\widetilde{K}_{s_2,T}^2 \text{vec}\left(C_{t_2-s_2,T}^{'}(u)\, C_{t_2-s_2,T}(u)\right)^{'} \widetilde{C}\text{vec}\left(C_{t_1-s_1,T}^{'}(u)\, C_{t_1-s_1,T}(u)\right) \\
&= \sum_{t_1=2}^{T}\sum_{t_2=2}^{T}\sum_{\substack{s_1,s_2=1\\ s_1\neq s_2}}^{(t_1-1)\wedge(t_2-1)} \widetilde{K}_{t_1,T}^2\widetilde{K}_{t_2,T}^2\widetilde{K}_{s_1,T}^2\widetilde{K}_{s_2,T}^2 \text{tr}\left(C_{t_1-s_1,T}^{'}(u)\, C_{t_1-s_1,T}(u)\right)\text{tr}\left(C_{t_2-s_2,T}^{'}(u)\, C_{t_2-s_2,T}(u)\right) \\
&= \mathbb{E}\left(Z_1\right)^2 - \sum_{t_1=2}^{T}\sum_{t_2=2}^{T}\sum_{s=1}^{(t_1-1)\wedge(t_2-1)} \widetilde{K}_{t_1,T}^2\widetilde{K}_{t_2,T}^2\widetilde{K}_{s,T}^4 \text{tr}\left(C_{t_1-s,T}^{'}(u)\, C_{t_1-s,T}(u)\right)\text{tr}\left(C_{t_2-s,T}^{'}(u)\, C_{t_2-s,T}(u)\right) \\
&\equiv \mathbb{E}\left(Z_1\right)^2 - RSZ_2,
\end{aligned}$$

where RSZ_2 can be further given by

$$\begin{aligned}
RSZ_2 = &\sum_{t_1=2}^{T}\sum_{t_2=2}^{T}\sum_{s=1}^{(t_1-1)\wedge(t_2-1)} \widetilde{K}_{t_1,T}^2\widetilde{K}_{t_2,T}^2\widetilde{K}_{s,T}^4 \left(\frac{h}{4\pi^2T^2M}\sum_{j,k=1}^{M}\text{tr}\left(\theta_j^{'}\theta_k\right)\cos\left((t_1-s)\lambda_j\right)\cos\left((t_1-s)\lambda_k\right)\right) \\
&\times \left(\frac{h}{4\pi^2T^2M}\sum_{j,k=1}^{M}\text{tr}\left(\theta_j^{'}\theta_k\right)\cos\left((t_2-s)\lambda_j\right)\cos\left((t_2-s)\lambda_k\right)\right).
\end{aligned}$$

Note that by Nielsen (2005, pp. 294), $\|\theta_j\| = O\left(\left(\frac{M}{j}\right)^{d_{X_a}(u)+d_\varepsilon(u)-2\delta(u)}\right)$ as θ_j in our study is numerically identical to the one in Nielsen (2005), and thus using this property $|RSZ_2|$ can be bounded by

$$\begin{aligned}
&|RSZ_2| \\
&\leq \frac{1}{T}\sum_{t_2=1}^{T}\widetilde{K}_{t_2,T}^2\frac{h}{4\pi^2M}\frac{1}{T}\sum_{j,k=1}^{M}\left|\text{tr}\left(\theta_j^{\prime}\theta_k\right)\right| \\
&\times \left|\sum_{t_1=1}^{T}\sum_{s=1}^{T}\widetilde{K}_{t_1,T}^2\widetilde{K}_{s,T}^4\frac{h}{4\pi^2T^2M}\sum_{j,k=1}^{M}\text{tr}\left(\theta_j^{\prime}\theta_k\right)\cos\left((t_1-s)\lambda_j\right)\cos\left((t_1-s)\lambda_k\right)\right|
\end{aligned}$$

$$= O(Mh)\left|\sum_{t_1=1}^{T}\sum_{s=1}^{T}\widetilde{K}^2_{t_1,T}\widetilde{K}^4_{s,T}\frac{h}{4\pi^2T^3M}\sum_{j,k=1}^{M}\operatorname{tr}\left(\theta_j'\theta_k\right)\cos\left(\left(t_1-s\right)\lambda_j\right)\cos\left(\left(t_1-s\right)\lambda_k\right)\right|$$

$$\leq O(Mh)\frac{1}{4\pi^2M}\sum_{j,k=1}^{M}\left|\operatorname{tr}\left(\theta_j'\theta_k\right)\right|\left|\frac{h}{T^3}\sum_{t_1=1}^{T}\sum_{s=1}^{T}\widetilde{K}^2_{t_1,T}\widetilde{K}^4_{s,T}\cos\left(\left(t_1-s\right)\lambda_j\right)\cos\left(\left(t_1-s\right)\lambda_k\right)\right|$$

$$= O(Mh)\frac{1}{4\pi^2M}\sum_{j,k=1}^{M}\left|\operatorname{tr}\left(\theta_j'\theta_k\right)\right|$$

$$\times\left|\frac{h}{T^3}\sum_{t_1=1}^{T}\sum_{s=1}^{T}\widetilde{K}^2_{t_1,T}\widetilde{K}^4_{s,T}\frac{1}{4}\left(e^{i(t_1-s)\lambda_{j+k}}+e^{-i(t_1-s)\lambda_{j-k}}+e^{i(t_1-s)\lambda_{j-k}}+e^{-i(t_1-s)\lambda_{j+k}}\right)\right|$$

$$\equiv O(Mh)\frac{1}{4\pi^2M}\sum_{j,k=1}^{M}\left|\operatorname{tr}\left(\theta_j'\theta_k\right)\right|\left|\frac{h}{T^3}\sum_{t_1=1}^{T}\sum_{s=1}^{T}\widetilde{K}^2_{t_1,T}\widetilde{K}^4_{s,T}\frac{1}{4}C(\lambda)\right|,$$

where in $C(\lambda)$ anyone of the four terms can lead to the same order, so it is sufficient to analyze one of them. We take $e^{i(t_1-s)\lambda_{j+k}}$ for instance, and have that

$$\frac{1}{4\pi^2M}\sum_{j,k=1}^{M}\left|tr\left(\theta_j^{'}\theta_k\right)\right|\left|\frac{h}{T^3}\sum_{t_1=1}^{T}\sum_{s=1}^{T}\widetilde{K}^2_{t_1,T}\widetilde{K}^4_{s,T}\frac{1}{4}e^{i(t_1-s)\lambda_{j+k}}\right|$$

$$\leq\frac{1}{4\pi^2M}\sum_{j,k=1}^{M}\left|tr\left(\theta_j^{'}\theta_k\right)\right|\left|\frac{\sqrt{h}}{T}\sum_{t_1=1}^{T}\widetilde{K}^2_{t_1,T}e^{it_1\lambda_{j+k}}\right|\frac{\sqrt{h}}{T^2}\sum_{s=1}^{T}\widetilde{K}^4_{s,T}=O\left(\frac{1}{T\sqrt{h}}\right)$$

using the same reasoning as in (A.2.23), and the definition that $\widetilde{K}^2_{t,T}=\frac{K^2_{h,tu}}{\frac{1}{T}\sum_{t=1}^{T}K^2_{h,tu}}$. Therefore $|RSZ_2|=O\left(\frac{Mh}{T\sqrt{h}}\right)=o(1)$. Next by Assumption 1(i) $\mathbb{E}\left(\zeta_s\zeta_s^{'}\otimes\zeta_s\zeta_s^{'}\right)=\widetilde{B}$, then SZ_1 follows that

$$SZ_1\leq\left\|\widetilde{B}\right\|\sum_{t_1=2}^{T}\sum_{t_2=2}^{T}\sum_{s=1}^{(t_1-1)\wedge(t_2-1)}\widetilde{K}^2_{t_1,T}\widetilde{K}^2_{t_2,T}\widetilde{K}^4_{s,T}\left\|C^{'}_{t_2-s,T}(u)\,C_{t_2-s,T}(u)\right\|\left\|C^{'}_{t_1-s,T}(u)\,C_{t_1-s,T}(u)\right\|,$$

which has the same order as $|Z_{21,b}|$ in (A.2.20) and thus negligible. Therefore we can conclude that $\mathbb{E}\left(Z_1^2\right)=\mathbb{E}\left(Z_1\right)^2+o(1)$ and thus the variance of Z_1 is negligible, which then implies that

$$Z_1=\sum_{a=1}^{p}\sum_{b=1}^{p}\eta_a\eta_b\Theta^*\left(\frac{G^{\frac{1}{2}}_{X_{aa}}G^{\frac{1}{2}}_{X_{bb}}G_\varepsilon}{\left(1-d_{X_a}(u)-d_\varepsilon(u)+2\delta(u)\right)}\right)+o_p(1).$$

Then we complete the proof of arguemnt (i).

(ii) As in Lobato (1999), it is sufficient to prove $\sum_{t=1}^{T}\mathbb{E}\left(Z^4_{t,T}(u)\right)\to 0$. Note that

$$\sum_{t=1}^{T}\mathbb{E}\left(Z^4_{t,T}(u)\right)$$

$$=\sum_{t=2}^{T}\mathbb{E}\left(\sum_{s<t}\sum_{r<t}\widetilde{\zeta}_s^{'}C'_{t-s,T}(u)\,\widetilde{\zeta}_t\widetilde{\zeta}_t^{'}C_{t-r,T}(u)\,\widetilde{\zeta}_r\right)^2$$

$$=\sum_{t=2}^{T}\mathbb{E}\left[\sum_{s<t}\sum_{r<t}\widetilde{\zeta}_s^{'}C'_{t-s,T}(u)\,\widetilde{\zeta}_t\widetilde{\zeta}_t^{'}C_{t-r,T}(u)\,\widetilde{\zeta}_r\sum_{v<t}\sum_{w<t}\widetilde{\zeta}_v^{'}C'_{t-v,T}(u)\,\widetilde{\zeta}_t\widetilde{\zeta}_t^{'}C_{t-w,T}(u)\,\widetilde{\zeta}_w\right]$$

$$
\begin{aligned}
&= \sum_{t=2}^{T} \sum_{s,r,v,w<t} \mathbb{E}\left\{ \operatorname{tr}\left[C_{t-w,T}(u)\, \widetilde{\zeta}_w \widetilde{\zeta}_s' C_{t-s,T}'(u)\, \widetilde{\zeta}_t \widetilde{\zeta}_t' C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{t-v,T}'(u)\, \widetilde{\zeta}_t \widetilde{\zeta}_t' \right] \right\} \\
&= \sum_{t=2}^{T} \widetilde{K}_{t,T}^4 \sum_{s,r,v,w<t} \mathbb{E}\left[\operatorname{vec}\left(C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{t-v,T}'(u) \right)' \left(\widetilde{\zeta}_t \widetilde{\zeta}_t' \otimes \widetilde{\zeta}_t \widetilde{\zeta}_t' \right) \operatorname{vec}\left(C_{t-s,T}(u)\, \widetilde{\zeta}_s \widetilde{\zeta}_w' C_{t-w,T}'(u) \right) \right] \\
&\leq \left\| \widetilde{B} \right\| \sum_{t=2}^{T} \widetilde{K}_{t,T}^4 \sum_{s,r,v,w<t} \mathbb{E} \left\| \operatorname{vec}\left(C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{t-v,T}'(u) \right) \right\| \left\| \operatorname{vec}\left(C_{t-s,T}(u)\, \widetilde{\zeta}_s \widetilde{\zeta}_w' C_{t-w,T}'(u) \right) \right\| \\
&= \left\| \widetilde{B} \right\| \sum_{t=2}^{T} \widetilde{K}_{t,T}^4 \sum_{s,r,v,w<t} \mathbb{E} \left\| C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{t-v,T}'(u) \right\| \left\| C_{t-s,T}(u)\, \widetilde{\zeta}_s \widetilde{\zeta}_w' C_{t-w,T}'(u) \right\| \\
&\leq \left\| \widetilde{B} \right\| \sum_{t=2}^{T} \widetilde{K}_{t,T}^4 \sum_{r,v<t} \left(\mathbb{E} \left\| C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{t-v,T}'(u) \right\|^2 \right)^{\frac{1}{2}} \sum_{s,w<t} \left(\mathbb{E} \left\| C_{t-s,T}(u)\, \widetilde{\zeta}_s \widetilde{\zeta}_w' C_{t-w,T}'(u) \right\|^2 \right)^{\frac{1}{2}}.
\end{aligned}
$$

Note that by the same reasoning as how we derive (A.2.19), we have

$$
\begin{aligned}
&\mathbb{E} \left\| C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' C_{t-v,T}'(u) \right\|^2 \\
&= \mathbb{E}\left[\operatorname{tr}\left(C_{t-v,T}'(u)\, C_{t-v,T}(u)\, \widetilde{\zeta}_v \widetilde{\zeta}_r' C_{t-r,T}'(u)\, C_{t-r,T}(u)\, \widetilde{\zeta}_r \widetilde{\zeta}_v' \right) \right] \\
&\leq R \widetilde{K}_{v,T}^2 \widetilde{K}_{r,T}^2 \left\| C_{t-v,T}'(u)\, C_{t-v,T}(u) \right\| \left\| C_{t-r,T}'(u)\, C_{t-r,T}(u) \right\|
\end{aligned}
$$

with some constant R, where the last inequality holds by Assumption 1(i) and by the derivation of (A.2.25). Therefore it is enough to consider EF_2 as it is dominant in order when summing over t, v and r. Then by substitution,

$$
\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}\left(Z_{t,T}^4(u) \right) &\leq R \sum_{t=2}^{T} \widetilde{K}_{t,T}^4 \sum_{r,v<t} \widetilde{K}_{v,T} \widetilde{K}_{r,T} \sqrt{ \left\| C_{t-v,T}'(u)\, C_{t-v,T}(u) \right\| \left\| C_{t-r,T}'(u)\, C_{t-r,T}(u) \right\| } \\
&\quad \times \sum_{s,w<t} \widetilde{K}_{s,T} \widetilde{K}_{w,T} \sqrt{ \left\| C_{t-s,T}'(u)\, C_{t-s,T}(u) \right\| \left\| C_{t-w,T}'(u)\, C_{t-w,T}(u) \right\| } \\
&= R \sum_{t=2}^{T} \widetilde{K}_{t,T}^4 \left(\sum_{r<t} \widetilde{K}_{r,T} \sqrt{ \left\| C_{t-r,T}'(u)\, C_{t-r,T}(u) \right\| } \right)^2 \left(\sum_{s<t} \widetilde{K}_{s,T} \sqrt{ \left\| C_{t-s,T}'(u)\, C_{t-s,T}(u) \right\| } \right)^2 \\
&\leq R \sum_{t=2}^{T} \widetilde{K}_{t,T}^4 \left(\sum_{r<t} \widetilde{K}_{r,T}^2 \right)^2 \left(\sum_{s<t} \left\| C_{t-s,T}(u) \right\|^4 \right)^2 \\
&\leq R h^{-2} \left(\sum_{r=1}^{T} \widetilde{K}_{r,T}^2 \right)^2 \sum_{t=1}^{T} \left(\sum_{s<t} \left\| C_{t-s,T}(u) \right\|^4 \right)^2 \\
&= O\left(T^2 h^{-2} \right) \left(\sum_{t=1}^{T-1} (T-t) \left\| C_{t-s,T}(u) \right\|^4 \right)^2 \\
&\leq O\left(T^2 h^{-2} \right) \left(T \sum_{t=1}^{T-1} \left\| C_{t-s,T}(u) \right\|^4 \right)^2 = O\left(\frac{M^2 h^2}{T^2} \right),
\end{aligned}
$$

where the last equality holds because

$$\sum_{t=1}^{T-1} \left\| C_{t-s,T}\left(u\right) \right\|^{4} = O\left(\sum_{t=1}^{\lfloor T/M \rfloor} \frac{M^{2}h^{2}}{T^{4}} + \sum_{\lfloor T/M \rfloor +1}^{T} \frac{h^{2}}{t^{4}M^{2}} \right)$$

$$= O\left(\frac{Mh^{2}}{T^{3}} + \frac{h^{2}}{T^{3}M^{2}} \right) = O\left(\frac{Mh^{2}}{T^{3}} \right),$$

then we can conclude that $\sum_{t=1}^{T} \mathbb{E}\left(Z_{t,T}^{4}\left(u\right)\right) = o\left(1\right)$.

This completes the proof of Proposition A.1.5. ■

Appendix B

Technical Results for Chapter 3

B.1 Proofs of the Main Results

In this appendix we prove the main results in the paper. The proofs call upon some technical lemmas that proved in the online supplement. Throughout the proof, we use C to denote some absolute constant that may vary across lines. We use $A \lesssim B$ to denote $A/B = O_p(1)$.

B.1.1 Proofs of the Results in Section 3.3

To prove Theorem 3.3.1, we need the following four lemmas.

Lemma B.1.1 *Suppose Assumptions A–E and the other conditions for Theorem 3.3.1 hold. Let* $H = \left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right)V_{NT}^{-1}$ *and* $\delta_{NT} = \min\left(N^{1/2}, T^{1-\max(2d_\varepsilon,1/2)}\right)$. *Then*

$$\frac{1}{T}\left\|\hat{F} - F^0 H\right\|^2 = O_p\left(\left\|\hat{\beta} - \beta^0\right\|^2 + \delta_{NT}^{-2}\right).$$

Lemma B.1.2 *Suppose Assumptions A–E and the other conditions for Theorem 3.3.1 hold. We have*

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{\varepsilon_i'\hat{F}}{T} = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}} + N^{-1} + N^{-\frac{1}{2}}\left\|\hat{\beta} - \beta^0\right\|\right).$$

Lemma B.1.3 *Suppose Assumptions A–E and the other conditions for Theorem 3.3.1 hold. Let* $J_8 = -\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\varepsilon_k\varepsilon_k'\hat{F}G\lambda_i$, *where* $G = \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}$. *Then*

$$J_8 = A_{NT}^o + O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\beta} - \beta^0\right\| + \delta_{NT}^{-1}\right)\right),$$

where $A_{NT}^o = -\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\Omega_k\hat{F}G\lambda_i = O_p\left(T^{2d_\varepsilon-1}\right)$ *and* $\Omega_k = E\left(\varepsilon_k\varepsilon_k'\right)$ *for every* $k = 1, \ldots, N$.

Lemma B.1.4 *Suppose Assumptions A–E and the other conditions for Theorem 3.3.1 hold. We have*

$$\frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{\hat{F}}\right]\varepsilon_i$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{F^0} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{F^0}\right]\varepsilon_i + C_{NT}^o + o_p\left(\left\|\hat{\beta}-\beta^0\right\|\right) + O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\delta_{NT}^{-1} + N^{-\frac{1}{2}}T^{d_\varepsilon-1}\right),$$

where $C_{NT}^o = \frac{1}{NT}\sum_{i=1}^{N}\frac{(X_i-V_i)'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{N}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\varepsilon_i = O_p\left(\frac{1}{N}\right).$

Proof of Theorem 3.3.1. By the definition of $\hat{\beta}$, we have

$$\left(\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}X_i\right)\left(\hat{\beta}-\beta^0\right) = \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}F^0\lambda_i + \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\varepsilon_i. \tag{B.1.1}$$

For the first term on the right hand side (rhs), we have $\mathbf{M}_{\hat{F}}F^0 = \mathbf{M}_{\hat{F}}(F^0 - \hat{F}H^{-1})$ where H is defined in Lemma B.1.1 as $H = \left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right)V_{NT}^{-1} \equiv G^{-1}V_{NT}^{-1}$. The asymptotic invertibility of H and V_{NT} can be proved as in Proposition 1 of Bai (2009), as its proof does not involve any premise of serial persistence, and it holds under long range dependence as well. Then

$$\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}F^0\lambda_i = \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left(F^0-\hat{F}H^{-1}\right)\lambda_i = \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left(F^0-\hat{F}V_{NT}G\right)\lambda_i.$$

Let $\hat{\delta} = \hat{\beta}-\beta^0$. By the eigenvalue problem in (3.2.6),

$$\begin{aligned}\hat{F}V_{NT} &= \frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\hat{\delta}'X_i'\hat{F} - \frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\lambda_i'F^{0\prime}\hat{F} - \frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\varepsilon_i'\hat{F}\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i\hat{\delta}'X_i'\hat{F} - \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\hat{\delta}'X_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i\varepsilon_i'\hat{F}\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\lambda_i'F^{0\prime}\hat{F} + \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i\lambda_i'F^{0\prime}\hat{F} \qquad \text{(B.1.2)}\\ &\equiv I_1+\cdots+I_9.\end{aligned}$$

Then $\hat{F}V_{NT}G - F^0 = (I_1+\cdots+I_8)\,G$ and

$$\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}F^0\lambda_i = -\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left[I_1+\cdots+I_8\right]G\lambda_i \equiv J_1+\cdots+J_8.$$

We can derive the order of J_ℓ's by using the same reasoning as used in the proof of Lemma

B.1.1. For J_1, we have

$$\begin{aligned}\|J_1\| &= \left\| \frac{1}{NT}\sum_{i=1}^{N} X_i'\mathbf{M}_{\hat{F}} \frac{1}{NT}\sum_{k=1}^{N} X_k\hat{\delta}\hat{\delta}'X_k'\hat{F}G\lambda_i \right\| \\ &\leq \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|X_i\|\|\lambda_i\| \frac{1}{NT}\sum_{k=1}^{N}\|X_k\|^2 \left\|\hat{\delta}\right\|^2 \frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\|G\| \\ &\lesssim \left\|\hat{\delta}\right\|^2 \left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}} \left(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right)^{\frac{1}{2}} \frac{1}{NT}\sum_{k=1}^{N}\|X_k\|^2 = O_p\left(\left\|\hat{\delta}\right\|^2\right),\end{aligned}$$

where we use the fact that $\|\mathbf{M}_{\hat{F}}\|_{\text{sp}} = 1$, $\frac{1}{\sqrt{T}}\left\|\hat{F}\right\| = \sqrt{R}$, and Assumption B(i) and B(iv). For J_3 we have

$$\begin{aligned}J_3 &= \frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\frac{X_i'\mathbf{M}_{\hat{F}}X_k}{T}\right)\left(\frac{\varepsilon_k'F^0H}{T}\right)G\lambda_i\hat{\delta} \\ &+ \frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{X_i'\mathbf{M}_{\hat{F}}X_k}{T}\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}G\lambda_i\hat{\delta} \equiv J_{31}+J_{32}.\end{aligned}$$

Note that

$$\begin{aligned}\|J_{31}\| &= \left\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\frac{X_i'\mathbf{M}_{\hat{F}}X_k}{T}\right)\left(\frac{\varepsilon_k'F^0H}{T}\right)G\lambda_i\hat{\delta}\right\| \\ &\lesssim \left\|\hat{\delta}\right\| \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|X_i\|\|\lambda_i\| \left\|\frac{\varepsilon_k'F^0H}{T}\right\| \frac{1}{N\sqrt{T}}\sum_{k=1}^{N}\|X_k\| \\ &\leq \left\|\hat{\delta}\right\| \frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2 \left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}} \left[\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'F^0H}{T}\right\|^2\right]^{\frac{1}{2}} \\ &\lesssim \left\|\hat{\delta}\right\| \left[\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'F^0H}{T}\right\|^2\right]^{\frac{1}{2}} \equiv \left\|\hat{\delta}\right\| \bar{J}_{31},\end{aligned}$$

by Assumption B(i) and B(iv). By arguments as used to show (B.2.2) in the proof of Lemma B.1.2 and Assumption D(i), we can show that $\bar{J}_{31}^2 = O_p\left(T^{2d_\varepsilon-1}\right) = o_p(1)$. Therefore $\|J_{31}\| = o_p\left(\left\|\hat{\delta}\right\|\right)$. Similarly, we can show $\|J_{32}\| = o_p\left(\left\|\hat{\delta}\right\|\right)$ by adopting Lemma B.1.1. Then $\|J_3\| = o_p\left(\left\|\hat{\delta}\right\|\right)$. By the same token, we can show that

$$\|J_5\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left(\frac{1}{NT}\sum_{k=1}^{N}\varepsilon_k\hat{\delta}'X_k'\hat{F}\right)G\lambda_i\right\| = o_p\left(\left\|\hat{\delta}\right\|\right).$$

For J_4, we have

$$\|J_4\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left(\frac{1}{NT}\sum_{k=1}^{N}F^0\lambda_k\hat{\delta}'X_k'\hat{F}\right)G\lambda_i\right\|$$

$$
\begin{aligned}
&= \frac{1}{N^2T^2} \left\| \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \left(F^0 - \hat{F}H^{-1} \right) \sum_{k=1}^{N} \lambda_k \hat{\delta}' X_k' \hat{F} G \lambda_i \right\| \\
&\lesssim \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|X_i\|^2 \right\} \left\{ \frac{1}{N} \sum_{i=1}^{N} \|\lambda_i\|^2 \right\}^{1/2} \frac{\left\| F^0 - \hat{F}H^{-1} \right\|}{\sqrt{T}} \left\| \hat{\delta} \right\| = o_p\left(\left\| \hat{\delta} \right\| \right)
\end{aligned}
$$

by Lemma B.1.1. For J_6, we have

$$
\begin{aligned}
\|J_6\| &= \left\| \frac{1}{N^2T} \sum_{i=1}^{N} \sum_{k=1}^{N} X_i' \mathbf{M}_{\hat{F}} F \lambda_k \frac{\varepsilon_k' \hat{F}}{T} G \lambda_i \right\| \\
&= \left\| \frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \left(F^0 - \hat{F}H^{-1} \right) \left(\frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{\varepsilon_k' \hat{F}}{T} \right) G \lambda_i \right\| \\
&\lesssim \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} \|X_i\| \, \|\lambda_i\| \left\| \frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{\varepsilon_k' \hat{F}}{T} \right\| \frac{1}{\sqrt{T}} \left\| F^0 - \hat{F}H^{-1} \right\| \\
&\lesssim \left\| \frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{\varepsilon_k' \hat{F}}{T} \right\| \frac{1}{\sqrt{T}} \left\| F^0 - \hat{F}H^{-1} \right\| \\
&= O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left(1 + \left\| \hat{\delta} \right\| \right) \right) O_p\left(\left\| \hat{\delta} \right\| + \delta_{NT}^{-1} \right) \\
&= o_p\left(\left\| \hat{\delta} \right\| \right) + O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \delta_{NT}^{-1} \right)
\end{aligned}
$$

by Lemmas B.1.1 and B.1.2.

As in Bai (2009), J_2 and J_7 directly enter the asymptotic distribution and J_8 contributes to the bias under possible long range dependence. For J_8, we make the following decomposition:

$$
\begin{aligned}
J_8 &= -\frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \varepsilon_k \varepsilon_k' \hat{F} G \lambda_i \\
&= -\frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \Omega_k \hat{F} G \lambda_i - \frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \left(\varepsilon_k \varepsilon_k' - \Omega_k \right) F^0 H G \lambda_i \\
&\quad - \frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \left(\varepsilon_k \varepsilon_k' - \Omega_k \right) \left(\hat{F} - F^0 H \right) G \lambda_i \equiv J_{81} + J_{82} + J_{83}.
\end{aligned}
$$

By Lemma B.1.3, $J_{81} = A_{NT}^o = O_p\left(T^{2d_\varepsilon - 1} \right)$ and

$$
J_{82} + J_{83} = O_p\left(N^{-\frac{1}{2}} T^{2d_\varepsilon - 1} + N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left(\left\| \hat{\delta} \right\| + \delta_{NT}^{-1} \right) \right) = o_p\left(\left\| \hat{\delta} \right\| \right) + O_p\left(N^{-1} T^{d_\varepsilon - \frac{1}{2}} + N^{-\frac{1}{2}} T^{3d_\varepsilon - \frac{3}{2}} \right).
$$

In sum, we have

$$
\frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} F^0 \lambda_i = J_2 + J_7 + A_{NT}^o + o_p\left(\left\| \hat{\delta} \right\| \right) + O_p\left(N^{-1} T^{d_\varepsilon - \frac{1}{2}} + N^{-\frac{1}{2}} T^{3d_\varepsilon - \frac{3}{2}} \right). \quad \text{(B.1.3)}
$$

Combining (B.1.1) and (B.1.3) yields

$$\left(\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}X_i+o_p(1)\right)\hat{\delta}-J_2=\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\varepsilon_i+J_7+A_{NT}^o+O_p\left(N^{-1}T^{d_\varepsilon-\frac{1}{2}}+N^{-\frac{1}{2}}T^{3d_\varepsilon-\frac{3}{2}}\right),$$

which then implies that

$$\begin{aligned}&\left[D_{NT}\left(\hat{F}\right)+o_p(1)\right]\left(\hat{\beta}-\beta^0\right)\\&=\frac{1}{NT}\sum_{i=1}^{N}\left(X_i'\mathbf{M}_{\hat{F}}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{\hat{F}}\right)\varepsilon_i+A_{NT}^o+O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon-1}+N^{-\frac{1}{2}}T^{3d_\varepsilon-\frac{3}{2}}\right).\end{aligned}\tag{B.1.4}$$

Then by Lemma B.1.4 and the conditions that $\frac{T}{N}\to\rho>0$ and $d_{Z,\max}>d_\varepsilon$,

$$\begin{aligned}&\left[D\left(\hat{F}\right)+o_p(1)\right]\rho_{NT}\left(\hat{\beta}-\beta^0\right)\\&=\frac{\rho_{NT}}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{\hat{F}}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{\hat{F}}\right]\varepsilon_i+\rho_{NT}A_{NT}^o+\rho_{NT}O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon-1}+N^{-\frac{1}{2}}T^{3d_\varepsilon-\frac{3}{2}}\right)\\&=\frac{\rho_{NT}}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{F^0}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{F^0}\right]\varepsilon_i+\rho_{NT}\left(A_{NT}^o+C_{NT}^o\right)+\rho_{NT}O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon-1}+N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\delta_{NT}^{-1}\right),\\&=\frac{\rho_{NT}}{NT}\sum_{i=1}^{N}Z_i'\varepsilon_i+\rho_{NT}\left(A_{NT}^o+C_{NT}^o\right)+o_p(1),\end{aligned}$$

where recall that $Z_i=\mathbf{M}_{F^0}X_i-\frac{1}{N}\sum_{k=1}^{N}a_{ik}\mathbf{M}_{F^0}X_k$. By Assumption E(i), $D_{NT}\left(F^0\right)=\frac{1}{NT}\sum_{i=1}^{N}Z_i'Z_i\xrightarrow{p}D_0>0$. Using this assumption and Lemma B.1.1, we can readily show that $D_{NT}(\hat{F})=D_{NT}\left(F^0\right)+O_p\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)=D_0+o_p(1)$. It follows that

$$\rho_{NT}\left(\hat{\beta}-\beta^0-\frac{1}{T^{1-2d_\varepsilon}}A_{NT}^o-\frac{1}{N}C_{NT}^o\right)\xrightarrow{d}\mathcal{N}\left(0,D_0^{-1}\mathbf{\Sigma}D_0^{-1}\right),$$

where

$$\begin{aligned}A_{NT}&=-D_{NT}\left(F^0\right)^{-1}\frac{1}{NT^{1+2d_\varepsilon}}\sum_{i=1}^{N}X_i'\mathbf{M}_F\frac{1}{N}\sum_{k=1}^{N}\Omega_k\hat{F}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,\ \text{and}\\C_{NT}&=-D_{NT}\left(F^0\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\frac{\left(X_i-V_i\right)'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\varepsilon_i.\end{aligned}$$

This completes the proof of Theorem 3.3.1. ■

B.1.2 Proofs of the Results in Section 3.4

To prove Proposition 3.4.1, we need the following lemma whose proofs can be found in Supplemental Material.

Lemma B.1.5 *Under Assumptions A–D and* A^**–*B^* *we have*

(i) $\sup_{\tilde{W}_F \in \mathcal{W}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{X,i}^{*} \mathbf{M}_{\tilde{W}_F} W_{\varepsilon,i}^{*} \right\| = o_p(1)$;

(ii) $\sup_{\tilde{W}_F \in \mathcal{W}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i' W_{F^0}^{*} \mathbf{M}_{\tilde{W}_F} W_{\varepsilon,i} \right\| = o_p(1)$;

(iii) $\sup_{\tilde{W}_F \in \mathcal{W}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{\varepsilon,i}^{*} \left(\mathbf{P}_{\tilde{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}} \right) W_{\varepsilon,i}^{*} \right\| = o_p(1)$.

Proof of Proposition 3.4.1. The proof of this proposition follows closely to the proof of Proposition 1 in Bai (2009, pp. 1264) using the modified theory of consistency of an extremum estimator. Let $\delta = \beta - \beta^0$. By definition, the FDLS estimator $\left(\tilde{\beta}, \hat{W}_F\right)$ solves the following concentrated minimization problem as

$$\left(\tilde{\beta}, \hat{W}_F\right) = \arg \min_{\beta \in \mathbb{R}^P, \tilde{W}_F \in \mathcal{W}} S_{NT}\left(\beta, \tilde{W}_F\right),$$

where $\mathcal{W} = \left\{ \tilde{W}_F \in \mathbb{C}^{L \times R} : \tilde{W}_F = W_F \tilde{\Gamma}_F, \tilde{W}_F^{*} \tilde{W}_F / T = \mathbb{I}_R \right\}$. Recall the original objective function is given by (3.4.3) and (3.4.4) as

$$\begin{aligned} SSR(\beta, W_F, \Lambda) &= \sum_{i=1}^{N} (W_{Y,i} - W_{X,i}\beta - W_F\lambda_i)^{*} (W_{Y,i} - W_{X,i}\beta - W_F\lambda_i) \\ &= \sum_{i=1}^{N} \left(W_{Y,i} - W_{X,i}\beta - \tilde{W}_F\tilde{\lambda}_i\right)^{*} \left(W_{Y,i} - W_{X,i}\beta - \tilde{W}_F\tilde{\lambda}_i\right). \end{aligned}$$

Let $W_{U,i} = W_{U,i}(\beta) = W_{Y,i} - W_{X,i}\beta$. As in (3.4.7), we concentrate $\tilde{\lambda}_i$ out by plugging

$$\tilde{\lambda}_i = \left(\tilde{W}_F^{*}\tilde{W}_F\right)^{-1} \tilde{W}_F^{*} (W_{Y,i} - W_{X,i}\beta) = \tilde{W}_F^{*} (W_{Y,i} - W_{X,i}\beta)/T \equiv \tilde{W}_F^{*} W_{U,i}/T$$

into the above objective function and then simplify to obtain the concentrated objective function:

$$\begin{aligned} S_{NT}\left(\beta, \tilde{W}_F\right) &= \frac{1}{NT} \sum_{i=1}^{N} \left(W_{Y,i} - W_{X,i}\beta - \tilde{W}_F\tilde{\lambda}_i\right)^{*} \left(W_{Y,i} - W_{X,i}\beta - \tilde{W}_F\tilde{\lambda}_i\right) - \frac{1}{NT} \sum_{i=1}^{N} W_{\varepsilon,i}^{*} \mathbf{M}_{\tilde{W}_F} W_{\varepsilon,i} \\ &= \frac{1}{NT} \sum_{i=1}^{N} \left(W_{U,i} - \tilde{W}_F\tilde{W}_F^{*} W_{U,i}/T\right)^{*} \left(W_{U,i} - \tilde{W}_F\tilde{W}_F^{*} W_{U,i}/T\right) - \frac{1}{NT} \sum_{i=1}^{N} W_{\varepsilon,i}^{*} \mathbf{M}_{\tilde{W}_F} W_{\varepsilon,i} \\ &= \frac{1}{NT} \sum_{i=1}^{N} (W_{Y,i} - W_{X,i}\beta)^{*} \mathbf{M}_{\tilde{W}_F} (W_{Y,i} - W_{X,i}\beta) - \frac{1}{NT} \sum_{i=1}^{N} W_{\varepsilon,i}^{*} \mathbf{M}_{\tilde{W}_F} W_{\varepsilon,i}. \end{aligned}$$

As in Bai (2009), we approximate $S_{NT}\left(\beta, \tilde{W}_F\right)$ with another random function $\tilde{S}_{NT}\left(\beta, \tilde{W}_F\right)$

as follows

$$S_{NT}\left(\beta,\tilde{W}_F\right)=\tilde{S}_{NT}\left(\delta,\tilde{W}_F\right)+\delta'\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i}+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{X,i}\delta$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\lambda_i'W_{F^0}^{*}\mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i}^{*}+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{F^0}\lambda_i$$
$$+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^{*}\left(\mathbf{P}_{\tilde{W}_F}-\mathbf{P}_{\tilde{W}_{F^0}}\right)W_{\varepsilon,i},$$

where

$$\tilde{S}_{NT}\left(\beta,\tilde{W}_F\right)=\delta'\left(\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{X,i}\right)\delta+\operatorname{tr}\left[\left(\frac{W_{F^0}^{*}\mathbf{M}_{\tilde{W}_F}W_{F^0}}{T}\right)\left(\frac{\Lambda'\Lambda}{N}\right)\right]$$
$$+\delta'\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{F^0}\lambda_i+\frac{1}{NT}\sum_{i=1}^{N}\lambda_i'W_{F^0}^{*}\mathbf{M}_{\tilde{W}_F}W_{X,i}\delta.$$

where $\delta=\beta-\beta^0$. By Lemma B.1.5, $S_{NT}\left(\beta,\tilde{W}_F\right)=\widetilde{S}_{NT}\left(\beta,\tilde{W}_F\right)+o_p(1)$ uniformly over $\beta\in\mathbb{R}^P$ and $\tilde{W}_F\in\mathcal{W}$. Then we can focus on the approximated objective function $\widetilde{S}_{NT}\left(\beta,\tilde{W}_F\right)$. Note that $\widetilde{S}_{NT}\left(\beta^0,HW_{F^0}\right)=0$ for any asymptotically invertible matrix H by construction, and because $\tilde{\Gamma}_F$ is also invertible, $\widetilde{S}_{NT}\left(\beta^0,H\tilde{W}_{F^0}\right)=0$ holds as well. Then analogously to the proof in Bai (2009), we denote

$$A=\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{X,i},\ B=\frac{1}{T}\left(\frac{\Lambda'\Lambda}{N}\otimes\mathbb{I}_L\right),\ C=\frac{1}{NT}\sum_{i=1}^{N}\left(\lambda_i\otimes\mathbf{M}_{\tilde{W}_F}W_{X,i}\right),$$

and $\eta=\operatorname{vec}\left(\mathbf{M}_{\tilde{W}_F}W_{F^0}\right)$, where $\operatorname{vec}(\cdot)$ is the vectorization operator that stack the columns of a matrix into a single column vector. Then

$$\begin{aligned}\widetilde{S}_{NT}\left(\beta,\tilde{W}_F\right)&=\delta'A\delta+\eta^*B\eta+\delta'C^*\eta+\eta^*C\delta\\&=\delta'\left(A-C^*B^{-1}C\right)\delta+\left(\eta^*+\delta'C^*B^{-1}\right)B\left(\eta+B^{-1}C\delta\right)\\&\equiv\delta'D^{\dagger}(\tilde{W}_F)\delta+\theta^*B\theta.\end{aligned}$$

By Assumption B(iv), B is positive definite asymptotically, and so as $D^{\dagger}(\tilde{W}_F)$ by Assumption B*(ii). Therefore $\tilde{S}_{NT}\left(\beta,\tilde{W}_F\right)>0$ if $\delta=\beta-\beta^0\neq0$ or $\tilde{W}_F\neq H\tilde{W}_{F^0}$, which implies $\left(\beta^0,H\tilde{W}_{F^0}\right)$ is the unique minimizer of $\tilde{S}_{NT}\left(\beta,\tilde{W}_F\right)$ over the restrictions. With this result, in conjunction with the uniform approximation before and arguments used in Bai (2009, pp. 1265), we can conclude that $\tilde{\beta}$ is a consistent estimator for β.

Next, for (ii), note that the proof in Bai (2009, pp. 1265) is extended directly to our frequency domain setup given the consistency of $\tilde{\beta}$. ■

To prove Theorem 3.4.2, we require the following lemmas that are proved in the Online Supplement.

Lemma B.1.6 *Suppose Assumption A, B, and A^*-E^*, and the other conditions of Theorem 3.4.2 hold. Let $\tilde{H} = \left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\left(\frac{\tilde{W}_{F^0}^*\hat{W}_F}{T}\right)V_{NL}^{-1}$. Then*

$$T^{-\frac{1}{2}}\left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\| = O_p\left(\delta_{W1,NT}\left\|\tilde{\beta}-\beta^0\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right),$$

where $\delta_{W1,NT} = \gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$.

Lemma B.1.7 *Suppose Assumptions A, B, and A^*-E^*, and the other conditions of Theorem 3.4.2 hold. We have*

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i\left(\frac{W_{\varepsilon,i}^*\hat{W}_F}{T}\right) = O_p\left(\delta_{W,NL}\left\|\tilde{\beta}-\beta^0\right\| + N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right)$$

where $\delta_{W,NL} = N^{-\frac{1}{2}}\gamma_L^{1-d_{X,\max}-d_\varepsilon}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$.

Lemma B.1.8 *Suppose Assumptions A, B, and A^*-E^*, and suppose the other conditions of Theorem 3.4.2 hold. Let*

$$\tilde{J}_8 = -\frac{1}{NT}\sum_{i=1}^{N} Re\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\right)$$

and

$$A_{NT} = -\frac{1}{NT}\sum_{i=1}^{N} Re\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\check{G}\lambda_i\right),$$

with $Diag\left(|W_{\varepsilon,kj}|^2\right)$ a diagonal matrix of $|W_{\varepsilon,kj}|^2$, $j = 1,\ldots,L$. Then

$$\tilde{J}_8 = A_{NT} + O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$$
$$+ O_p\left(\left(T^{2d_\varepsilon-1}\gamma_L^{d_{F,\min}-d_{X,\max}} + N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon+(d_{F,\min}-d_{X,\max})}\right)\left(\delta_{W1,NT}\left(\tilde{\beta}-\beta^0\right) + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)$$

and

$$A_{NT} = O_p\left(\frac{1}{L}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right).$$

Lemma B.1.9 *Suppose Assumptions A, B, and A^*-E^*, and suppose the other conditions of Theorem 3.4.2 hold. Recall that $W_{V,i} = \frac{1}{N}\sum_{k=1}^{N}a_{ik}W_{X,k}$. We have*

$$\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}Re\left[\left(W_{X,i}^*-W_{V,i}^*\right)\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right]$$

$$= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} Re\left[\left(W_{X,i}^* - W_{V,i}^*\right) \mathbf{M}_{W_{F^0}} W_{\varepsilon,i}\right] + o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta} - \beta^0\right)\right) + o_p(1).$$

Proof of Theorem 3.4.2. Recall that

$$\tilde{\beta} = \left[\frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{X,i}\right)\right]^{-1} \left[\frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{Y,i}\right)\right],$$

where $W_{Y,i} = W_{X,i}\beta^0 + \tilde{W}_{F^0}\tilde{\lambda}_i + W_{\varepsilon,i}$. Then

$$\begin{aligned}\tilde{\delta} \equiv \tilde{\beta} - \beta^0 = &\left[\frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{X,i}\right)\right]^{-1} \\ &\times \left[\frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \tilde{W}_{F^0}\tilde{\lambda}_i\right) + \frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{\varepsilon,i}\right)\right],\end{aligned}$$

which implies that

$$\left[\frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{X,i}\right)\right]\left(\tilde{\beta} - \beta^0\right) \tag{B.1.5}$$

$$= \frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \tilde{W}_{F^0}\tilde{\lambda}_i\right) + \frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{\varepsilon,i}\right). \tag{B.1.6}$$

First, we study $\frac{1}{NT}\sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{F^0}\lambda_i\right)$. Note that $\mathbf{M}_{\hat{W}_F}\tilde{W}_{F^0} = \mathbf{M}_{\hat{W}_F}\left(\tilde{W}_{F^0} - \hat{W}_F\tilde{H}^{-1}\right)$, where the asymptotic invertibility of $\tilde{H}$ can be proved using similar reasoning as used in the time domain. We consider the following eigenvalue problem

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(W_{Y,i} - W_{X,i}\tilde{\beta}\right)\left(W_{Y,i} - W_{X,i}\tilde{\beta}\right)^*\right]\hat{W}_F = \hat{W}_F V_{NL}.$$

By expanding $W_{Y,i}$ in the above equation, we have

$$\begin{aligned}\hat{W}_F V_{NL} = &\frac{1}{NT}\sum_{i=1}^{N} W_{X,i}\tilde{\delta}\tilde{\delta}' W_{X,i}^*\hat{W}_F - \frac{1}{NT}\sum_{i=1}^{N} W_{X,i}\tilde{\delta}\lambda_i' W_{F^0}^*\hat{W}_F - \frac{1}{NT}\sum_{i=1}^{N} W_{X,i}\tilde{\delta} W_{\varepsilon,i}^*\hat{W}_F^* \\ &- \frac{1}{NT}\sum_{i=1}^{N} W_{F^0}\lambda_i\tilde{\delta}' W_{X,i}^*\hat{W}_F - \frac{1}{NT}\sum_{i=1}^{N} W_{\varepsilon,i}\tilde{\delta}' W_{X,i}^*\hat{W}_F + \frac{1}{NT}\sum_{i=1}^{N} W_{F^0}\lambda_i W_{\varepsilon,i}^*\hat{W}_F \\ &+ \frac{1}{NT}\sum_{i=1}^{N} W_{\varepsilon,i}\lambda_i' W_{F^0}^*\hat{W}_F + \frac{1}{NT}\sum_{i=1}^{N} W_{\varepsilon,i} W_{\varepsilon,i}^*\hat{W}_F + \frac{1}{NT}\sum_{i=1}^{N} \tilde{W}_{F^0}\tilde{\lambda}_i\tilde{\lambda}_i'\tilde{W}_{F^0}^*\hat{W}_F \\ &\equiv \tilde{I}_1 + \cdots + \tilde{I}_9.\end{aligned} \tag{B.1.7}$$

This, in conjunction with the definition of $\tilde{H}$ given in the proof of Lemma B.1.6, it implies that

$$\tilde{W}_{F^0} - \hat{W}_F \tilde{H}^{-1} = -\left(\tilde{I}_1 + \cdots + \tilde{I}_8\right)\left(\tilde{W}_{F^0}^* \hat{W}_F / T\right)^{-1}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)^{-1}$$
$$= -\left(\tilde{I}_1 + \cdots + \tilde{I}_8\right)\tilde{G}.$$

Then

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \tilde{W}_{F^0}\tilde{\lambda}_i\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \tilde{W}_{F^0}\tilde{\Gamma}_F^{-1}\lambda_i\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F}\left(\tilde{W}_{F^0} - \hat{W}_F\tilde{H}^{-1}\right)\tilde{\Gamma}_F^{-1}\lambda_i\right) \\
&= -\frac{1}{NT}\sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F}\left(\tilde{I}_1 + \cdots + \tilde{I}_8\right)\left(\tilde{W}_{F^0}^*\hat{W}_F/T\right)^{-1}\tilde{\Gamma}_F\left(\Lambda'\Lambda/N\right)^{-1}\lambda_i\right) \\
&\equiv -\frac{1}{NT}\sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F}\left(\tilde{I}_1 + \cdots + \tilde{I}_8\right)\check{G}\lambda_i\right) \equiv \tilde{J}_1 + \cdots + \tilde{J}_8.
\end{aligned}$$

It is easy to show that $\tilde{H} = O_p\left(\gamma_L^{1-2d_{F,\max}}\right)$ and $\check{G} = O_p\left(\gamma_L^{d_{F,\min}-\frac{1}{2}}\right)$ by Assumption B(iv) and B*(iii). For $\tilde{J}_1$, we have

$$\begin{aligned}
\left\|\tilde{J}_1\right\| &= \left\|\frac{1}{NT}\sum_{i=1}^{N} W_{X,i}^* \mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N} W_{X,k}^*\tilde{\delta}\tilde{\delta}' W_{X,k}\hat{W}_F^*\right)\check{G}\lambda_i\right\| \\
&\lesssim \left\|\tilde{G}\right\|\left\|\tilde{\delta}\right\|^2 \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\left\|W_{X,i}\right\|\left\|\lambda_i\right\| \frac{1}{NT}\sum_{k=1}^{N}\left\|W_{X,k}\right\|^2 \\
&\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|^2\left\{\frac{1}{NT}\sum_{i=1}^{N}\left\|W_{X,i}\right\|^2\right\}^{3/2}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i\right\|^2\right]^{\frac{1}{2}} \\
&= \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|^2 O_p\left(\gamma_L^{3/2-3d_{X,\max}}\right) O_p(1) = O_p\left(\gamma_L^{1+d_{F,\min}-3d_{X,\max}}\left\|\tilde{\delta}\right\|^2\right),
\end{aligned}$$

where we use the additional facts that $\left\|\mathbf{M}_{\hat{W}_F}\right\|_{\text{sp}} = 1$ and $\frac{1}{NT}\sum_{k=1}^{N}\left\|W_{X,k}\right\|^2 = O_p\left(\gamma_L^{1-2d_{X,\max}}\right)$ by Assumption B*(i). Then we can express $\tilde{J}_1 = -\tilde{J}_1^*\tilde{\delta}$ with $\tilde{J}_1^* = O_p\left(\gamma_L^{1+d_{F,\min}-3d_{X,\max}}\tilde{\delta}\right)$. Next as in time domain, $\tilde{J}_2$ will enter the asymptotic distribution. For $\tilde{J}_3$, we make the following decomposition:

$$\begin{aligned}
\tilde{J}_3 &= \frac{-1}{NT}\sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N} W_{X,k}\tilde{\delta} W_{\varepsilon,k}^*\hat{W}_F\right)\check{G}\lambda_i\right) \\
&= \frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} \operatorname{Re}\left[\left(\frac{W_{X,i}^*\mathbf{M}_{\hat{W}_F} W_{X,k}}{T}\right)\left(\frac{W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H}}{T}\right)\check{G}\lambda_i\tilde{\delta}\right]
\end{aligned}$$

$$+\frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathrm{Re}\left[\left(\frac{W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}W_{X,k}}{T}\right)\left(\frac{W_{\varepsilon,k}^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)}{T}\right)\check{G}\lambda_i\tilde{\delta}\right]\equiv\tilde{J}_{3,1}+\tilde{J}_{3,2}.$$

First,

$$\begin{aligned}\left\|\tilde{J}_{3,1}\right\| &\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\left[\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{W_{\varepsilon,k}^{*}\tilde{W}_{F^0}\tilde{H}}{T}\right\|^2\right]^{\frac{1}{2}}\\
&\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\left[\frac{1}{NT^2}\sum_{k=1}^{N}\left\|W_{\varepsilon,k}^{*}\tilde{W}_{F^0}\right\|^2\right]^{\frac{1}{2}}\left\|\tilde{H}\right\|\\
&= \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|O_p\left(\gamma_L^{1-2d_{X,\max}}\right)O_p(1)O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p(1)O_p\left(\gamma_L^{1-2d_{F,\max}}\right)\\
&= O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{2-2d_{F,\max}-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\left\|\tilde{\delta}\right\|\right)\end{aligned}$$

by Assumption B(iv), B*(i), C*(i) and D*(iii). And using the same notation involving $\tilde{J}_1^*$, we have $\tilde{J}_{3,1}=-\tilde{J}_{3,1}^{*}\tilde{\delta}$ with $\tilde{J}_{3,1}^{*}=O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{2-2d_{F,\max}-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\right)$. And by the same reasoning and Lemma B.1.6,

$$\begin{aligned}\left\|\tilde{J}_{3,2}\right\| &\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\left[\frac{1}{NT}\sum_{k=1}^{N}\|W_{\varepsilon,k}\|^2\right]^{\frac{1}{2}}\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\\
&= O_p\left(\gamma_L^{1-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\left\|\tilde{\delta}\right\|\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right),\end{aligned}$$

where $\delta_{W1,NT}$ is defined in Lemma B.1.6, which altogether form the order of $\tilde{J}_3$. And we can write $\tilde{J}_{3,2}=-\tilde{J}_{3,2}^{*}\tilde{\delta}$ with

$$\tilde{J}_{3,2}^{*}=O_p\left(\gamma_L^{1-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)+O_p\left(\gamma_L^{1-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\delta_{W1,NT}\tilde{\delta}\right).$$

Next

$$\begin{aligned}\left\|\tilde{J}_4\right\| &= \left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N}W_{F^0}\lambda_k\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right)\check{G}\lambda_i\right\|\\
&= \left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N}\left(\tilde{W}_{F^0}-\hat{W}_F\tilde{H}^{-1}\right)\tilde{\lambda}_k\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right)\check{G}\lambda_i\right\|\\
&\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\left(\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\|W_{X,i}\|\,\|\lambda_i\|\right)^2\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F\right\|\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\left\|\tilde{H}^{-1}\tilde{\Gamma}_F^{-1}\right\|\\
&\lesssim \gamma_L^{2d_{F,\min}-1}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\\
&= O_p\left(\gamma_L^{2d_{F,\min}-1}\left\|\tilde{\delta}\right\|\right)O_p\left(\gamma_L^{1-2d_{X,\max}}\right)O_p(1)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right)\\
&= O_p\left(\gamma_L^{2d_{F,\min}-2d_{X,\max}}\delta_{W1,NT}\left\|\tilde{\delta}\right\|^2\right)+O_p\left(\gamma_L^{2d_{F,\min}-2d_{X,\max}}N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|\right),\end{aligned}$$

and thus we can write $\tilde{J}_4 = -\tilde{J}_4^* \tilde{\delta}$ with

$$\tilde{J}_4^* = O_p\left(\gamma_L^{2d_{F,\min}-2d_{X,\max}} N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon}\right) + O_p\left(\gamma_L^{2d_{F,\min}-2d_{X,\max}} \delta_{W1,NT} \tilde{\delta}\right).$$

Next,

$$\begin{aligned}
\left\|\tilde{J}_5\right\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{X,i}^* \mathbf{M}_{\hat{W}_F} \left(\frac{1}{NT} \sum_{k=1}^{N} W_{\varepsilon,k} \tilde{\delta}' W_{X,k}^* \hat{W}_F \right) \check{G} \lambda_i \right\| \\
&\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}} \left\|\tilde{\delta}\right\| \frac{1}{NT^{\frac{1}{2}}} \sum_{i=1}^{N} \|W_{X,i}\| \|\lambda_i\| \frac{1}{NT} \sum_{k=1}^{N} \|W_{\varepsilon,k}\| \|W_{X,k}\| \frac{1}{T^{\frac{1}{2}}} \left\|\hat{W}_F\right\| \\
&= O_p\left(\gamma_L^{d_{F,\min}-\frac{1}{2}} \left\|\tilde{\delta}\right\|\right) O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right) O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right) \\
&= O_p\left(\gamma_L^{1-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)} \left\|\tilde{\delta}\right\|\right),
\end{aligned}$$

and $\tilde{J}_5 = -\tilde{J}_5^* \tilde{\delta}$ with $\tilde{J}_5^* = O_p\left(\gamma_L^{1-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\right)$. Next for $\tilde{J}_6$, we have

$$\begin{aligned}
\left\|\tilde{J}_6\right\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{X,i}^* \mathbf{M}_{\hat{W}_F} \tilde{W}_{F^0} \tilde{\lambda}_k \left(\frac{1}{N} \sum_{k=1}^{N} \frac{W_{\varepsilon,k}^* \hat{W}_F}{T} \check{G} \lambda_i \right) \right\| \\
&= \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{X,i}^* \mathbf{M}_{\hat{W}_F} \left(\tilde{W}_{F^0} - \hat{W}_F \tilde{H}^{-1} \right) \tilde{\Gamma}_F^{-1} \left(\frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{W_{\varepsilon,k}^* \hat{W}_F}{T} \right) \check{G} \lambda_i \right\| \\
&\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}} \frac{1}{NT^{\frac{1}{2}}} \sum_{i=1}^{N} \|W_{X,i}\| \|\lambda_i\| \frac{1}{T^{\frac{1}{2}}} \left\| \hat{W}_F - \tilde{W}_{F^0} \tilde{H} \right\| \left\| \frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{W_{\varepsilon,k}^* \hat{W}_F}{T} \right\| \left\| \tilde{H}^{-1} \tilde{\Gamma}_F^{-1} \right\| \\
&= O_p\left(\gamma_L^{2d_{F,\min}-1}\right) O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right) O_p\left(\delta_{W1,NT} \left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon}\right) \\
&\times O_p\left(\delta_{W,NL} \left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right) \\
&= O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}+2d_{F,\min}-1} \delta_{W1,NT} \delta_{W,NL} \left\|\tilde{\delta}\right\|^2\right) \\
&+ O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}+2d_{F,\min}-1} \left(\delta_{W1,NT} N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} + \delta_{W,NL} N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon} \right) \left\|\tilde{\delta}\right\| \right) \\
&+ O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}+2d_{F,\min}-1} N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon}\right) \\
&\equiv O_p\left(\Delta_{1,NT} \left\|\tilde{\delta}\right\|^2 + \Delta_{2,NT} \left\|\tilde{\delta}\right\| + \Delta_{3,NT}\right),
\end{aligned}$$

by Lemma B.1.6 and B.1.7. Also we can write $\tilde{J}_6 = -\tilde{J}_{6,1}^* \tilde{\delta} + \tilde{J}_{6,2}^*$ with $\tilde{J}_{6,1}^* = O_p\left(\Delta_{1,NT} \tilde{\delta} + \Delta_{2,NT}\right)$, and $\tilde{J}_{6,2}^* = O_p\left(\Delta_{3,NT}\right)$. And same as $\tilde{J}_2$, $\tilde{J}_7$ contributes to the asymptotic distribution directly. Lastly for $\tilde{J}_8$, by Lemma B.1.8, we have $\tilde{J}_8 = A_{NT} + \check{J}_8$, where

$$\begin{aligned}
A_{NT} &= -\frac{1}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \frac{1}{NT} \sum_{k=1}^{N} Diag\left(|W_{\varepsilon,kj}|^2 \right) \hat{W}_F \check{G} \lambda_i \right) \\
&= O_p\left(\frac{1}{L} \gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon} \right), \qquad \text{(B.1.8)}
\end{aligned}$$

and

$$
\begin{aligned}
\left\|\check{J}_8\right\| &= O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right) \\
&\quad + O_p\left(\left(T^{2d_\varepsilon-1}\gamma_L^{d_{F,\min}-d_{X,\max}} + N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon+(d_{F,\min}-d_{X,\max})}\right)\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right) \\
&\equiv O_p\left(\check{\Delta}_{1,NT}\left\|\tilde{\delta}\right\| + \check{\Delta}_{2,NT}\right),
\end{aligned}
$$

and A_{NT} will enter the bias. As before we can write $\check{J}_8 = \check{J}^*_{8,1}\tilde{\delta} + \check{J}^*_{8,2}$ with $\check{J}^*_{8,1} = O_p\left(\check{\Delta}_{1,NT}\right)$ and $\check{J}^*_{8,2} = O_p\left(\check{\Delta}_{2,NT}\right)$.

Then summarizing all the results we have obtained so far, (B.1.5) and (B.1.6) can be written as

$$
\left[\frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,i}\right)\right]\left(\tilde{\beta}-\beta^0\right) = \left(\tilde{J}_1 + \cdots + \tilde{J}_8\right) + \frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right),
$$

which is equivalent to

$$
\begin{aligned}
&\left[\frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,i}\right) - \tilde{J}_2 + \left(\tilde{J}^*_1 + \tilde{J}^*_{3,1} + \tilde{J}^*_{3,2} + \tilde{J}^*_4 + \tilde{J}^*_5 + \tilde{J}^*_{6,1} + \check{J}^*_{8,1}\right)\right]\left(\tilde{\beta}-\beta^0\right) \\
&= \tilde{J}^*_{6,2} + A_{NT} + \check{J}_{8,2} + \frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right) + \tilde{J}_7. \qquad \text{(B.1.9)}
\end{aligned}
$$

By construction $D^\dagger_{NL}\left(\hat{W}_F\right) = \frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,i}\right) - \tilde{J}_2$, and denote $\tilde{J}_* = \tilde{J}^*_1 + \tilde{J}^*_{3,1} + \tilde{J}^*_{3,2} + \tilde{J}^*_4 + \tilde{J}^*_5 + \tilde{J}^*_{6,1}$ and $\hat{C}_{NL} = \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}W^*_{X,k}\mathbf{M}_{\hat{W}_F}\right)W_{\varepsilon,i}\right)$, then left-multiplying $\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z$ on both sides of (B.1.9) implies

$$
\gamma_L^{-1}\Gamma_Z\left[D^\dagger_{NL}\left(\hat{W}_F\right) + \tilde{J}_*\right]\Gamma_Z\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0\right) = \sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z\left(\tilde{J}^*_{6,2} + A_{NT} + \check{J}_{8,2}\right) + \hat{C}_{NL}.
$$

To proceed, it is easy to show $\gamma_L^{-1}\Gamma_Z\left[D^\dagger_{NL}\left(\hat{W}_F\right) - D^\dagger_{NL}\right]\Gamma_Z = o_p(1)$ using Lemma B.1.6 some other regularity conditions. We also need to show that $\gamma_L^{-1}\Gamma_Z\tilde{J}_*\Gamma_Z = o_p(1)$, which means $\tilde{J}_* = o_p\left(\gamma_L^{1-2d_{Z,\min}}\right)$ as the order is defined using matrix norm. Note that this argument can be proved because the following seven arguments hold under Assumption E*(ii) and E*(iii) about the relative magnitude among the memory parameters and the convergence rate of $\tilde{\delta}$:

$$
\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}^*_1\right\| = O_p\left(\gamma_L^{2d_{Z,\min}+d_{F,\min}-3d_{X,\max}}\left\|\tilde{\delta}\right\|\right) = o_p(1);
$$

$$
\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}^*_{3,1}\right\| = O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{1+2d_{Z,\min}-2d_{F,\max}-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\right) = o_p(1);
$$

$$
\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}^*_{3,2}\right\| = O_p\left(\gamma_L^{2d_{Z,\min}-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right) = o_p(1);
$$

$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_4^*\right\| = O_p\left(\gamma_L^{2d_{Z,\min}+2d_{F,\min}-2d_{X,\max}-1}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right) = o_p(1);$$

$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_5^*\right\| = O_p\left(\gamma_L^{2d_{Z,\min}-2d_{X,\max}+(d_{F,\min}-d_\varepsilon)}\left\|\tilde{\delta}\right\|\right) = o_p(1);$$

$$\begin{aligned}
&\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_{6,1}^*\right\| \\
&= \gamma_L^{2d_{Z,\min}-1}O_p\left(\Delta_{1,NT}\left\|\tilde{\delta}\right\| + \Delta_{2,NT}\right) \\
&= O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}+2d_{F,\min}-\frac{3}{2}}\delta_{W1,NT}\delta_{W,NL}\left\|\tilde{\delta}\right\|\right) \\
&+ O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}+2d_{F,\min}-\frac{3}{2}}\left(\delta_{W1,NT}N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} + \delta_{W,NL}N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right) \\
&= o_p(1);
\end{aligned}$$

$$\begin{aligned}
&\gamma_L^{2d_{Z,\min}-1}\left\|\check{J}_{8,1}^*\right\| \\
&= O_p\left(\delta_{W1,NT}\left(T^{2d_\varepsilon-1}\gamma_L^{d_{F,\min}-d_{X,\max}+2d_{Z,\min}-1} + N^{-\frac{1}{2}}\gamma_L^{2d_{Z,\min}-2d_\varepsilon+(d_{F,\min}-d_{X,\max})}\right)\right) \\
&= o_p(1)
\end{aligned}$$

Next in the following we examine the negligibility of $\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z\tilde{J}_{6,2}^*$ and $\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z\check{J}_{8,2}^*$. To be specific, as above we check whether $\tilde{J}_{6,2}^*$ and $\check{J}_{8,2}^*$ are $o_p\left(\frac{1}{\sqrt{NL}}\gamma_L^{1-d_{Z,\min}-d_\varepsilon}\right)$. And as a result,

$$\begin{aligned}
&\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\left\|\tilde{J}_{6,2}^*\right\| \\
&= O_p\left(\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\Delta_{3,NT}\right) \\
&= O_p\left(\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\gamma_L^{\frac{1}{2}-d_{X,\max}+2d_{F,\min}-1}N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right) \\
&+ O_p\left(N^{-\frac{1}{2}}\gamma_L^{1+2d_{F,\min}+d_{Z,\min}-3d_{F,\max}-d_{X,\max}-d_\varepsilon}\right) = o_p\left(1\right),
\end{aligned}$$

and

$$\begin{aligned}
&\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\left\|\check{J}_{8,2}^*\right\| \\
&= O_p\left(\gamma_L^{\frac{3}{2}+d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-d_\varepsilon}\right) \\
&+ O_p\left(\frac{L^{\frac{1}{2}}}{T^{1-2d_\varepsilon}}\gamma_L^{d_{F,\min}+d_{Z,\min}-d_{X,\max}-d_{F,\max}} + \gamma_L^{\frac{3}{2}-2d_\varepsilon+d_{Z,\min}+d_{F,\min}-d_{X,\max}-d_{F,\max}}\right) = o_p(1)
\end{aligned}$$

Lastly we analyze $\hat{C}_{NL}$. Since by Assumption E*(ii), $C_{NL} \xrightarrow{d} \mathcal{N}\left(0,\Sigma\right)$ with

$$C_{NL} = \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\left(W_{X,i}^*\mathbf{M}_{W_{F^0}} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}W_{X,k}^*\mathbf{M}_{W_{F^0}}\right)W_{\varepsilon,i}\right),$$

and by Lemma B.1.9 $\hat{C}_{NL} = C_{NL} + C^*_{NL}\tilde{\delta} + o_p(1)$, where $C^*_{NL} = o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\right)$. Combining all the terms we have so far implies that

$$
\begin{aligned}
&\left[\gamma_L^{-1}\Gamma_Z D^{\dagger}_{NL}\Gamma_Z + o_p(1)\right]\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0\right)\\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\left(W_{X,i}\mathbf{M}_{W^*_{F^0}} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}W_{X,k}\mathbf{M}_{W^*_{F^0}}\right)W^*_{\varepsilon,i}\right) + \sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z A_{NT} + o_p(1)\\
&\equiv C_{NL} + \tilde{A}^W_{NT} + o_p(1).
\end{aligned}
$$

The asymptotic bias term satisfies

$$
\begin{aligned}
\tilde{A}^W_{NT} &= O_p\left(\sqrt{\frac{N}{L}}\gamma_L^{1+d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-d_\varepsilon}\right)\\
&= \sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}O_p\left(\frac{1}{L}\gamma_L^{1+2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right),
\end{aligned}
\tag{B.1.10}
$$

which could be explosive. Note that C_{NL} satisfies the CLT as $C_{NL}\xrightarrow{d}\mathcal{N}(0,\Sigma)$ by Assumption D*(ii). Denoting $D^W_{NL} = \gamma_L^{-1}\Gamma_Z D^{\dagger}_{NL}\Gamma_Z$, we can rewrite the above formula as

$$
\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0\right) = \left(D^W_{NL}\right)^{-1}C_{NL} + \left(D^W_{NL}\right)^{-1}\tilde{A}^W_{NT} + o_p(1),
$$

because $\left(D^W_{NL}\right)^{-1} = O_p(1)$ by Assumption D(i), and thus using the right hand side of (B.1.10), we have

$$
\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0 - A^W_{NT}\right) = \left(D^W_{NL}\right)^{-1}C_{NL} + o_p(1),
$$

where

$$
\begin{aligned}
A^W_{NT} &= \Gamma_Z\left(D^W_{NL}\right)^{-1}\Gamma_Z\gamma_L^{-1}A_{NT}\\
&= -\Gamma_Z\left(D^W_{NL}\right)^{-1}\Gamma_Z\frac{\gamma_L^{-1}}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\check{G}\lambda_i\right),
\end{aligned}
$$

with $A^W_{NT} = O_p(\phi_L)$ and $\phi_L = \frac{\gamma_L^{2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}}{L}$, which is based on (B.1.8) and the definition of $Gamma_Z$.

Summarizing all we have so far, we complete the proof of Theorem 3.4.2. ■

Proof of Theorem 3.4.3. In the first stage we prove the consistency of asymptotic covariance estimator. By construction, it is equivalent to show the consistency of both

$\gamma_L^{-1}\Gamma_Z\hat{D}_{NL}^W\Gamma_Z$ and $\gamma_L^{2d_\varepsilon-1}\Gamma_Z\hat{\Sigma}_{NL}^W\Gamma_Z$. Firstly recall that

$$\begin{aligned}\hat{D}_{NL}^W &= \frac{1}{NT}\sum_{i=1}^N \text{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{X,i}\right) - \frac{1}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \text{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{X,k}\widehat{a}_{ik}\right)\right] \\ &= \frac{1}{NT}\sum_{i=1}^N \text{Re}\left(\hat{W}_{Z,i}^*\hat{W}_{Z,i}\right).\end{aligned}$$

Then by Assumption D*(i) it is sufficient to prove $\gamma_L^{-1}\Gamma_Z\left(\hat{D}_{NL}^W - D_{NL}^\dagger\right)\Gamma_Z = o_p(1)$, where by definition $D_{NL}^\dagger$ has the same form as $\tilde{D}$ only with $\hat{W}_F$ and $\widehat{a}_{ik}$ replaced by their true values. Then

$$\begin{aligned}\gamma_L^{-1}\Gamma_Z\left(\tilde{D} - D_{NL}^\dagger\right)\Gamma_Z &= \frac{\gamma_L^{-1}\Gamma_Z}{NT}\sum_{i=1}^N \text{Re}\left(W_{X,i}^*\left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{\tilde{W}_{F^0}}\right)W_{X,i}\right)\Gamma_Z \\ &\quad - \frac{\gamma_L^{-1}\Gamma_Z}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \text{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{X,k}\left(\widehat{a}_{ik} - a_{ik}\right)\right)\right]\Gamma_Z \\ &\quad - \frac{\gamma_L^{-1}\Gamma_Z}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \text{Re}\left(W_{X,i}^*\left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{\tilde{W}_{F^0}}\right)W_{X,k}^*a_{ik}\right)\right]\Gamma_Z \\ &= d_1 + d_2 + d_3.\end{aligned}$$

using the fact that $\mathbf{M}_{\tilde{W}_{F^0}} = \mathbf{M}_{W_{F^0}}$. Firstly as before, we denote $\tilde{\delta} = \tilde{\beta} - \beta$, and d_1 follows that

$$\begin{aligned}\|d_1\| &= \left\|\frac{\gamma_L^{-1}\Gamma_Z}{NT}\sum_{i=1}^N \text{Re}\left(W_{X,i}^*\left(\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right)W_{X,i}\right)\Gamma_Z\right\| \\ &\lesssim \gamma_L^{2d_{Z,\min}-1}\frac{1}{NT}\sum_{i=1}^N \|W_{X,i}\|^2\left\|\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right\| \\ &= O_p\left(\gamma_L^{\frac{1}{2}+2d_{Z,\min}-2d_{X,\max}-d_{F,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)^{\frac{1}{2}}\right) = o_p(1)\end{aligned}$$

by Assumption E*(iii) given the convergence rate of $\tilde{\beta}$, where the last two equalities holds by the following reasoning:

$$\begin{aligned}\left\|\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right\|^2 &= \text{tr}\left[\left(\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right)^2\right] = 2\text{tr}\left(I_R - \hat{W}_F^*\mathbf{P}_{\tilde{W}_{F^0}}\hat{W}_F/T\right) \\ &= 2\text{tr}\left(\mathbb{I}_R - \frac{\hat{W}_F^*\tilde{W}_{F^0}}{T}\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\frac{\tilde{W}_{F^0}^*\hat{W}_F}{T}\right),\end{aligned} \tag{B.1.11}$$

then using fact that

$$\frac{\tilde{W}_{F^0}^*\hat{W}_F}{T} = \frac{1}{T}\tilde{W}_{F^0}^*\tilde{W}_{F^0}\tilde{H} + \frac{1}{T}\tilde{W}_{F^0}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right),$$

where $\left\|\tilde{H}\right\| = O_p\left(\gamma_L^{1-2d_{F,\max}}\right)$ is defined as in Lemma B.1.6, we have

$$\begin{aligned}
&\frac{\hat{W}_F^*\tilde{W}_{F^0}}{T}\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\frac{\tilde{W}_{F^0}^*\hat{W}_F}{T} \\
&= \left[\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\tilde{H} + \frac{1}{T}\tilde{W}_{F^0}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\right]^*\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\left[\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\tilde{H} + \frac{1}{T}\tilde{W}_{F^0}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\right] \\
&= \tilde{H}^*\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)\tilde{H} + \frac{1}{T}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)^*\tilde{W}_{F^0}\tilde{H} + \frac{1}{T}\tilde{H}^*\tilde{W}_{F^0}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right) \\
&+ \frac{1}{T^2}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)^*\tilde{W}_{F^0}\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\tilde{W}_{F^0}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right) \equiv d_{11} + d_{12} + d_{13} + d_{14}.
\end{aligned}$$

Then for (B.1.11), we can analyze it through the decomposition above. First for $\operatorname{tr}(\mathbb{I}_R - d_{11})$ we have

$$\begin{aligned}
\operatorname{tr}(\mathbb{I}_R - d_{11}) &= \operatorname{tr}\left(\mathbb{I}_R - \tilde{H}^*\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)\tilde{H}\right) \\
&\leq \sqrt{R}\left\|\mathbb{I}_R - \tilde{H}^*\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)\tilde{H}\right\| = O_p\left(\gamma_L^{1-2d_{F,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)
\end{aligned}$$

by (B.2.10) in the proof of Lemma B.1.9. By the same reasoning we can deal with the Frobenius norm of d_{12}-d_{14} instead of their traces. To be specific, by Lemma B.1.6 and Assumption B*(iii), $\|d_{12}\|$ and $\|d_{13}\|$ have the same order as $\operatorname{tr}(\mathbb{I}_R - d_{11})$, which also dominate $\|d_{14}\|$. And by Assumption B(iv) d_3 have the same order as d_1.

Next for d_2, note that we can rewrite $a_{ik} = \tilde{\lambda}_k'\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{\lambda}_i$, and $\hat{a}_{ik} = \hat{\lambda}_k^*\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i$ with $\hat{\Lambda} = \overline{\hat{W}_U\hat{W}_F/T}$ where W_U is an $N \times L$ complex matrix such that $W_{U,i} = W_{Y,i} - W_{X,i}\beta$, thus $\hat{\lambda}_i = \hat{W}_F^*\hat{W}_{U,i}/T$. Since the conjugate transpose is compatible with the real vectors, we have

$$\begin{aligned}
\hat{a}_{ik} - a_{ik} &= \hat{\lambda}_k^*\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right) \\
&+ \hat{\lambda}_k^*\left[\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1} - \tilde{H}^*\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{H}\right]\tilde{H}^{-1}\tilde{\lambda}_i \\
&+ \left(\hat{\lambda}_k - \tilde{H}^{-1}\tilde{\lambda}_k\right)^*\tilde{H}^*\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{\lambda}_i \\
&\equiv b_{ik} + c_{ik} + d_{ik}
\end{aligned}$$

following the proof of Proposition 2 in Bai (2009), with $\tilde{H}$ defined as in Lemma B.1.6. Then d_2 can be decomposed into $d_{21} + d_{22} + d_{23}$ which correspond to b_{ik}, c_{ik} and d_{ik} respectively above. So it remains to study d_{21}, d_{22} and d_{23}, where d_{21} is firstly given by

$$\begin{aligned}
\|d_{21}\| &= \left\| \frac{\gamma_L^{-1}\Gamma_Z}{T} \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{k=1}^{N} \mathrm{Re} \left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{X,k} \hat{\lambda}_k^* \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} \left(\hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right) \right) \right] \Gamma_Z \right\| \\
&\lesssim \gamma_L^{2d_{Z,\min}-1} \frac{1}{NT^{\frac{1}{2}}} \left\| \sum_{i=1}^{N} \left(\hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right) W_{X,i}^* \right\| \frac{1}{NT^{\frac{1}{2}}} \sum_{k=1}^{N} \|W_{X,k}\| \left\| \hat{\lambda}_k^* \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} \right\| \\
&\lesssim \gamma_L^{2d_{Z,\min}-d_{X,\max}-\frac{1}{2}} \frac{1}{NT^{\frac{1}{2}}} \left\| \sum_{i=1}^{N} \left(\hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right) W_{X,i}^* \right\|.
\end{aligned}$$

by the fact that $\hat{\lambda}_k^* \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} \left(\hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right)$ is a scalar and the property of PCA in frequency domain. Then it is adequate to study the order of $\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right\|^2$. Since $\hat{\lambda}_i = \hat{W}_F^* \hat{W}_{U,i}/T$, we have

$$\begin{aligned}
\hat{\lambda}_i &= \hat{W}_F^* \left(W_{Y,i} - W_{X,i} \tilde{\beta} \right) / T \\
&= \frac{1}{T} \hat{W}_F^* \left(\tilde{W}_{F^0} \tilde{\lambda}_i + W_{\varepsilon,i} - W_{X,i} \tilde{\delta} \right) \\
&= \frac{1}{T} \hat{W}_F^* \hat{W}_F \tilde{H}^{-1} \tilde{\lambda}_i + \frac{1}{T} \hat{W}_F^* \left(\tilde{W}_{F^0} - \hat{W}_F \tilde{H}^{-1} \right) \tilde{\lambda}_i + \frac{1}{T} \hat{W}_F^* W_{\varepsilon,i} - \frac{1}{T} \hat{W}_F^* W_{X,i} \tilde{\delta} \\
&= \tilde{H}^{-1} \tilde{\lambda}_i + \frac{1}{T} \hat{W}_F^* \left(\tilde{W}_{F^0} - \hat{W}_F \tilde{H}^{-1} \right) \tilde{\lambda}_i + \frac{1}{T} \hat{W}_{F^0}^* W_{\varepsilon,i} - \frac{1}{T} \hat{W}_F^* W_{X,i} \tilde{\delta}
\end{aligned}$$

therefore by the same reasoning as we have used so far,

$$\begin{aligned}
&\frac{1}{NT^{\frac{1}{2}}} \left\| \sum_{i=1}^{N} \left(\hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right) W_{X,i}^* \right\| \\
&= O_p \left(\left(\gamma_L^{d_{F,\min}-d_{X,\max}} \delta_{W1,NT} + \gamma_L^{1-2d_{X,\max}} \right) \left\| \tilde{\delta} \right\| \right) \\
&+ O_p \left(N^{-\frac{1}{2}} \gamma_L^{1+d_{F,\min}-d_{F,\max}-d_{X,\max}-d_\varepsilon} + N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{1-d_{X,\max}-d_\varepsilon} \right)
\end{aligned} \tag{B.1.12}$$

by Lemma B.1.6 and D*(iii), which is $o_p(1)$ by convergence rate of $\tilde{\beta}$ and Assumption E*(iii). Therefore by substitution, $\|d_{21}\| = O_p \left(D_{1,NT} \left\| \tilde{\delta} \right\| + D_{2,NT} \right)$, with

$$D_{1,NT} = \gamma_L^{d_{F,\min}+2d_{Z,\min}-2d_{X,\max}-\frac{1}{2}} \delta_{W1,NT} + \gamma_L^{\frac{1}{2}+2d_{Z,\min}-3d_{X,\max}},$$

and

$$D_{2,NT} = N^{-\frac{1}{2}} \gamma_L^{\frac{1}{2}+d_{F,\min}+2d_{Z,\min}-2d_{X,\max}-d_{F,\max}-d_\varepsilon} + N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{1}{2}+2d_{Z,\min}-2d_{X,\max}-d_\varepsilon},$$

which implies the negligibility of $\|d_{21}\|$ by convergence rate of $\tilde{\beta}$ and Assumption E*(iii). And it is easy to see the same reasoning and order hold for d_{23}. Then it remains to analyze

d_{22}, which follows that

$$
\begin{aligned}
&\|d_{22}\| \\
&= \left\| \frac{\gamma_L^{-1}\Gamma_Z}{T} \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{k=1}^{N} \mathrm{Re} \left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{X,k} \hat{\lambda}_k^* \left[\left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} - \tilde{H}^* \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right)^{-1} \tilde{H} \right] \tilde{H}^{-1} \tilde{\lambda}_i \right) \right] \Gamma_Z \right\| \\
&\lesssim \gamma_L^{2d_{Z,\min}-1} \frac{1}{NT^{\frac{1}{2}}} \sum_{i=1}^{N} \|W_{X,i}\| \left\| \tilde{H}^{-1} \tilde{\lambda}_i \right\| \frac{1}{NT^{\frac{1}{2}}} \sum_{k=1}^{N} \|W_{X,k}\| \left\| \hat{\lambda}_k \right\| \left\| \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} - \tilde{H}^* \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right)^{-1} \tilde{H} \right\| \\
&= O_p \left(\gamma_L^{2d_{F,\min}+2d_{Z,\min}-2d_{X,\max}-1} \right) \left\| \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} - \tilde{H}^* \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right)^{-1} \tilde{H} \right\|.
\end{aligned}
$$

Since

$$
\begin{aligned}
&\left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} - \tilde{H}^* \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right)^{-1} \tilde{H} \\
&= \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} \left(\tilde{H}^{-1} \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right) \tilde{H}^{*-1} - \frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right) \tilde{H}^* \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right)^{-1} \tilde{H} \\
&= - \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} \left[\frac{1}{N} \left(\hat{\Lambda}^* - \tilde{H}^{-1} \tilde{\Lambda}^* \right) \tilde{\Lambda} \tilde{H}^{*-1} + \frac{1}{N} \hat{\Lambda}^* \left(\hat{\Lambda} - \tilde{\Lambda} \tilde{H}^{*-1} \right) \right] \\
&\times \tilde{H}^* \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right)^{-1} \tilde{H},
\end{aligned}
$$

whose order can be derived in a similar way from (B.1.12). To be specific,

$$
\begin{aligned}
\left\| \left(\frac{\hat{\Lambda}^* \hat{\Lambda}}{N} \right)^{-1} - \tilde{H}^* \left(\frac{\tilde{\Lambda}' \tilde{\Lambda}}{N} \right)^{-1} \tilde{H} \right\| &\lesssim \gamma_L^{\frac{5}{2}+d_{F,\min}-6d_{F,\max}} \frac{1}{\sqrt{N}} \left\| \hat{\Lambda}^* - \tilde{H}^{-1} \tilde{\Lambda}^* \right\| \\
&\leq \gamma_L^{\frac{5}{2}+d_{F,\min}-6d_{F,\max}} \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right\|^2 \right)^{\frac{1}{2}},
\end{aligned}
$$

and analogous to (B.1.13), we have

$$
\begin{aligned}
\left(\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\lambda}_i - \tilde{H}^{-1} \tilde{\lambda}_i \right\|^2 \right)^{\frac{1}{2}} &= O_p \left(\left(\gamma_L^{d_{F,\min}-\frac{1}{2}} \delta_{W1,NT} + \gamma_L^{\frac{1}{2}-d_{X,\max}} \right) \left\| \tilde{\delta} \right\| \right) \\
&+ O_p \left(N^{-\frac{1}{2}} \gamma_L^{\frac{1}{2}+d_{F,\min}-d_{F,\max}-d_\varepsilon} + L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} \right).
\end{aligned}
\tag{B.1.13}
$$

Therefore d_{22} is dominated by d_{21} and thus negligible as well. So we prove the consistency of $\hat{D}_{NL}^{W}$.

Next to show the consistency of Σ_{NL}^{W}, recall that

$$\hat{\Sigma}_{NL}^{W} = \frac{1}{N^2T^2}\sum_{i=1}^{N}\mathrm{Re}\left(\hat{W}_{Z,i}^{*}\hat{W}_{\varepsilon,i}\right)\mathrm{Re}\left(\hat{W}_{\varepsilon,i}^{*}\hat{W}_{Z,i}\right)$$

By Assumption D*(ii), it is sufficient to prove $NL\gamma_L^{2d_\varepsilon-2}\Gamma_Z\hat{\Sigma}_{NL}^{W}\Gamma_Z - \Sigma^\dagger = o_p(1)$, where

$$\Sigma^\dagger = \frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\Gamma_Z W_{Z,i}^{*}W_{\varepsilon,i}\right)\mathrm{Re}\left(W_{\varepsilon,i}^{*}W_{Z,i}\Gamma_Z\right)$$

under the additional condition of cross-sectional independence of $\zeta_{\varepsilon,i,t}$. To begin, we have the following decomposition:

$$\begin{aligned}
&NL\gamma_L^{2d_\varepsilon-2}\Gamma_Z\hat{\Sigma}_{NL}^{W}\Gamma_Z - \Sigma^\dagger \\
&= \frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\Gamma_Z\left[\mathrm{Re}\left(\hat{W}_{Z,i}^{*}\hat{W}_{\varepsilon,i}\right)\mathrm{Re}\left(\hat{W}_{\varepsilon,i}^{*}\hat{W}_{Z,i}\right) - \mathrm{Re}\left(W_{Z,i}^{*}W_{\varepsilon,i}\right)\mathrm{Re}\left(W_{\varepsilon,i}^{*}W_{Z,i}\right)\right]\Gamma_Z \\
&= \frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\Gamma_Z\mathrm{Re}\left(\hat{W}_{Z,i}^{*}\hat{W}_{\varepsilon,i} - W_{Z,i}^{*}W_{\varepsilon,i}\right)\mathrm{Re}\left(W_{\varepsilon,i}^{*}W_{Z,i}\right)\Gamma_Z \\
&+ \frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\Gamma_Z\mathrm{Re}\left(\hat{W}_{Z,i}^{*}\hat{W}_{\varepsilon,i}\right)\mathrm{Re}\left(\hat{W}_{\varepsilon,i}^{*}\hat{W}_{Z,i} - W_{\varepsilon,i}^{*}W_{Z,i}\right)\Gamma_Z \\
&\equiv w_1 + w_2.
\end{aligned}$$

Firstly for w_1, it follows that

$$\|w_1\| \le \left(\frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\left\|\Gamma_Z\left(\hat{W}_{Z,i}^{*}\hat{W}_{\varepsilon,i} - W_{Z,i}^{*}W_{\varepsilon,i}\right)\right\|^2\right)^{\frac{1}{2}}\left(\frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\left\|\Gamma_Z W_{Z,i}^{*}W_{\varepsilon,i}\right\|^2\right)^{\frac{1}{2}} \equiv \sqrt{w_{11}}\sqrt{w_{12}},$$

where by Assumption D*(iii) $w_{12} = O_p(1)$. And for w_{11}, we can firstly bound it by

$$w_{11} \le \frac{2\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\left\|\Gamma_Z\hat{W}_{Z,i}^{*}\left(\hat{W}_{\varepsilon,i} - W_{\varepsilon,i}\right)\right\|^2 + \frac{2\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\left\|\Gamma_Z\left(\hat{W}_{Z,i}^{*} - W_{Z,i}^{*}\right)W_{\varepsilon,i}\right\|^2 \equiv w_{11,1} + w_{11,2}.$$

By what we have analyzed the consistency of D_{NL}^{W}, it is easy to see the negligibility of $w_{11,2}$. And for $w_{11,1}$, note that

$$\begin{aligned}
\hat{W}_{\varepsilon,i} &= W_{Y,i} - W_{X,i}\tilde{\beta} - \hat{W}_F\hat{\lambda}_i \\
&= W_{\varepsilon,i} - W_{X,i}\tilde{\delta} + \tilde{W}_{F^0}\tilde{H}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right) + \left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^{-1}\tilde{\lambda}_i + \left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right),
\end{aligned} \tag{B.1.14}$$

where clearly the last term above is dominated by the others. Then $w_{11,1}$ follows that

$$
\begin{aligned}
w_{11,1} &\lesssim \frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\left\|\Gamma_Z W_{Z,i}^{*}W_{X,i}\tilde{\delta}\right\|^2+\frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\left\|\Gamma_Z W_{Z,i}^{*}\tilde{W}_{F^0}\tilde{H}\left(\hat{\lambda}_i-\tilde{H}^{-1}\tilde{\lambda}_i\right)\right\|^2\\
&+\frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\left\|\Gamma_Z W_{Z,i}^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^{-1}\tilde{\lambda}_i\right\|^2\\
&=o_p\left(\gamma_L^{1-2d_{X,\max}}\right)+o_p\left(\gamma_L^{1-2d_{F,\max}}\right)+O_p\left(\gamma_L^{2d_\varepsilon+2d_{F,\max}-1}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|^2\right)\\
&=o_p(1)
\end{aligned}
$$

by Assumption E*(iii) and (B.1.12), thus $w_1=o_p(1)$. Also by the same reasoning we have $w_2=o_p\left(1\right)$, which completes the proof of $NL\gamma_L^{2d_\varepsilon-2}\Gamma_Z\hat{\Sigma}_{NL}^{W}\Gamma_Z-\Sigma^{\dagger}=o_p\left(1\right)$. Combining the results we have so far it can be concluded that

$$
\left(\hat{\Sigma}_{NL}^{W}\right)^{-\frac{1}{2}}\left(\hat{D}_{NL}^{W}\right)\xrightarrow{p}\left(\Sigma_0^{W}\right)^{-\frac{1}{2}}\left(D_0^{W}\right)\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1},
$$

which illustrates the idea of self-normalization.

In the next stage we prove the validity of our bias correction. To be exact, we are about to prove

$$
\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\hat{A}_{NT}^{W}-A_{NT}^{W}\right)=o_p(1).
$$

Recall that

$$
\begin{aligned}
&\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\hat{A}_{NT}^{W}\\
&=-\left[\gamma_L^{-1}\Gamma_Z\hat{D}_{NL}^{W}\Gamma_Z\right]^{-1}\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}\hat{\Omega}_k\hat{W}_F\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}A_{NT}^{W}\\
&=-\left[\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z\right]^{-1}\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\tilde{G}\tilde{\lambda}_i\right)\\
&=-\left[\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z\right]^{-1}\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\tilde{G}\tilde{H}\tilde{H}^{-1}\tilde{\lambda}_i\right)\\
&=-\left[\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z\right]^{-1}\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\tilde{G}V_{NL}^{-1}\tilde{\lambda}_i\right).
\end{aligned}
$$

Note that the denominator parts above are $\gamma_L^{-1}\Gamma_Z\hat{D}_{NL}^{W}\Gamma_Z$ and $\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z$, and the negligibility of their difference has already been proved just now. So it is sufficient to consider the difference between the numerator parts, which by the same reasoning as before, can be decomposed by

$$\begin{aligned}&\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left[W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}\left(\hat{\Omega}_k-Diag\left(|W_{\varepsilon,kj}|^2\right)\right)\hat{W}_F V_{NL}^{-1}\hat{\lambda}_i\right]\\&+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F V_{NL}^{-1}\left(\hat{\lambda}_i-\tilde{H}^{-1}\tilde{\lambda}_i\right)\right)\\&\equiv D_1+D_2\end{aligned}$$

where we use the fact that $\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}=V_{NL}^{-1}$ by (3.4.7) and (3.4.10). For D_1, following the same reasoning as before,

$$\|D_1\|\lesssim\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-d_{X,\max}-\frac{1}{2}}\left\|\frac{1}{NT}\sum_{k=1}^{N}\left(\hat{\Omega}_k-Diag\left(|W_{\varepsilon,kj}|^2\right)\right)\right\|.$$

Since by (B.1.14),

$$\left(\left|\hat{W}_{\varepsilon,kj}\right|^2-|W_{\varepsilon,kj}|^2\right)\lesssim\left\|W_{X,kj}'\tilde{\delta}\right\|^2+\left\|\tilde{W}_{F^0,j}'\tilde{H}\left(\hat{\lambda}_k-\tilde{H}^{-1}\tilde{\lambda}_k\right)\right\|^2+\left\|\left(\hat{W}_{F,j}'-\tilde{W}_{F^0,j}'\tilde{H}\right)\tilde{H}^{-1}\tilde{\lambda}_k\right\|^2,$$

then $\|D_1\|$ will further follow that

$$\begin{aligned}&\|D_1\|\\&\lesssim\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-d_{X,\max}-\frac{1}{2}}\frac{1}{NT}\sum_{k=1}^{N}\sum_{j=1}^{L}\left|\left|\hat{W}_{\varepsilon,kj}\right|^2-|W_{\varepsilon,kj}|^2\right|\\&\leq\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-d_{X,\max}-\frac{1}{2}}\frac{1}{NT}\sum_{k=1}^{N}\left(\left\|W_{X,k}\tilde{\delta}\right\|^2+\left\|\tilde{W}_{F^0}\tilde{H}\left(\hat{\lambda}_k-\tilde{H}^{-1}\tilde{\lambda}_k\right)\right\|^2+\left\|\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^{-1}\tilde{\lambda}_k\right\|^2\right)\\&=O_p\left(\sqrt{NL}\gamma_L^{\frac{1}{2}+d_{Z,\min}+d_\varepsilon-3d_{X,\max}}\left\|\tilde{\delta}\right\|^2\right)+O_p\left(\sqrt{NL}\gamma_L^{\frac{1}{2}+d_{Z,\min}+d_\varepsilon-d_{X,\max}-d_{F,\max}}\right).\end{aligned}$$

And for $\|D_2\|$, we have

$$\|D_2\|\lesssim\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-d_{X,\max}-\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\lambda}_i-\tilde{H}^{-1}\tilde{\lambda}_i\right\|^2\right)^{\frac{1}{2}}\left\|\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\right\|,$$

It is easy to show the negligibility of both D_1 and D_2 by the same reasoning as before and Assumption C*(i) and E*(ii). So we complete the proof of Theorem 3.4.3. ■

B.1.3 Proofs of the Results in Section 3.5

Proof of Theorem 3.5.1. First we prove the consistency of $\tilde{\beta}_{(R_{\max})}$, which extends the proof of Theorem 4.1 in Moon and Weidner (2015) into frequency domain. To be specific,

$$\tilde{\beta}_{(R_{\max})} = \arg\min_{\beta \in \mathbb{R}^P} \mathcal{L}_{NT}^{R_{\max}}(\beta),$$

where by extending the objective function (3.4.3), and by considering (3.4.6), (3.4.7) and (3.4.8) together with the property of PCA,

$$\begin{aligned}
\mathcal{L}_{NT}^{R_{\max}}(\beta) &= \min_{\tilde{\Lambda} \in \mathbb{C}^{N \times R_{\max}}, \tilde{W}_F \in \mathbb{C}^{L \times R_{\max}}} \frac{1}{NT} \left\| W_Y - W_X \cdot \beta - \tilde{\Lambda} \tilde{W}_F' \right\|^2 \\
&= \min_{\tilde{W}_F \in \mathbb{C}^{L \times R_{\max}}} \frac{1}{NT} \text{tr}\left[(W_Y - W_X \cdot \beta) \mathbf{M}_{\tilde{W}_F} (W_Y - W_X \cdot \beta)^* \right] \\
&= \frac{1}{NT} \sum_{r=R_{\max}+1}^{L} \mu_r \left[(W_Y - W_X \cdot \beta)^* (W_Y - W_X \cdot \beta) \right]
\end{aligned}$$

subject to the identification restrictions, where $\mu_r(\cdot)$ represents the r-th largest eigenvalue. Note that $W_Y = W_X \cdot \beta^0 + \tilde{\Lambda}^0 \tilde{W}_{F^0}' + W_\varepsilon$, where we put the superscripts to emphasize the true values. Then

$$\begin{aligned}
&\mathcal{L}_{NT}^{R_{\max}}(\beta) \\
&= \min_{\tilde{\Lambda} \in \mathbb{C}^{N \times R_{\max}}, \tilde{W}_F \in \mathbb{C}^{L \times R_{\max}}} \frac{1}{NT} \left\| W_X \cdot \delta + W_\varepsilon + \tilde{\Lambda}^0 \tilde{W}_{F^0}' - \tilde{\Lambda} \tilde{W}_F' \right\|^2 \\
&\geq \min_{\tilde{\Lambda} \in \mathbb{C}^{N \times R_{\max}+R^0}, \tilde{W}_F \in \mathbb{C}^{L \times R_{\max}+R^0}} \frac{1}{NT} \left\| W_X \cdot \delta + W_\varepsilon - \tilde{\Lambda} \tilde{W}_F' \right\|^2 \\
&= \min_{\tilde{W}_F \in \mathbb{C}^{L \times R_{\max}+R^0}} \frac{1}{NT} \text{tr}\left[(W_X \cdot \delta + W_\varepsilon) \mathbf{M}_{\tilde{W}_F} (W_X \cdot \delta + W_\varepsilon)^* \right] \\
&= \min_{\tilde{W}_F \in \mathbb{C}^{L \times R_{\max}+R^0}} \frac{1}{NT} \text{tr}\left[(W_X \cdot \delta) \mathbf{M}_{\tilde{W}_F} (W_X \cdot \delta)^* \right] - \text{tr}\left(W_\varepsilon \mathbf{P}_{\tilde{W}_F} W_\varepsilon^* \right) - 2\text{tr}\left[(W_X \cdot \delta) \mathbf{P}_{\tilde{W}_F} W_\varepsilon^* \right] \\
&+ \text{tr}\left(W_\varepsilon W_\varepsilon^* \right) + 2\text{tr}\left[(W_X \cdot \delta) W_\varepsilon^* \right] \\
&\geq \frac{1}{NT} \sum_{r=R_{\max}+R^0+1}^{L} \mu_r \left[(W_X \cdot \delta)^* (W_X \cdot \delta) \right] + \frac{1}{NT} \text{tr}\left(W_\varepsilon W_\varepsilon^* \right) + \frac{1}{NT} 2\text{tr}\left[(W_X \cdot \delta) W_\varepsilon^* \right] \\
&- \frac{1}{NT} \left(2 \left(R_{\max} + R^0 \right) \left\| W_\varepsilon \right\|^2 + 2 \left(R_{\max} + R^0 \right) \left\| W_\varepsilon \right\| \left\| W_X \cdot \delta \right\| \right) \\
&\geq b \left\| \delta \right\|^2 + \frac{1}{NT} \text{tr}\left(W_\varepsilon W_\varepsilon^* \right) + O_p\left(\gamma_L^{1-2d_\varepsilon} \right) + O_p\left(\left\| \delta \right\| \gamma_L^{1-d_{X,\max}-d_\varepsilon} \right)
\end{aligned} \tag{B.1.15}$$

by Assumption B*(i), C*(i) and J, where $\delta = \beta - \beta^0$ following our notation before. Next it is easy to see

$$\mathcal{L}_{NT}^{R_{\max}}(\beta) \leq \mathcal{L}_{NT}^{R_{\max}}\left(\beta^0\right)$$

$$= \min_{\tilde{\Lambda}\in\mathbb{C}^{N\times R_{\max}},\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}}} \frac{1}{NT}\left\|W_\varepsilon + \tilde{\Lambda}^0\tilde{W}'_{F^0} - \tilde{\Lambda}\tilde{W}'_F\right\|^2$$
$$\leq \frac{1}{NT}\|W_\varepsilon\|^2 = \frac{1}{NT}\mathrm{tr}\left(W_\varepsilon W_\varepsilon^*\right). \qquad \text{(B.1.16)}$$

Then combining (B.1.15) and (B.1.16) we have

$$b\|\delta\|^2 + O_p\left(\gamma_L^{1-2d_\varepsilon}\right) + O_p\left(\|\delta\|\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right) \leq 0,$$

which implies a loose order of $\|\delta\|$ as $\left\|\tilde{\beta}_{(R_{\max})}-\beta\right\| = O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right)$ by Assumption E*(ii) and validates the consistency of $\tilde{\beta}_{(R_{\max})}$.

Next we try to prove of the consistency of ER estimator, which will closely follow the proof of Theorem 1 in AH. To proceed, we can complete the proof by showing the following three statements:

$$\frac{\tilde{\mu}_{NT,j}}{\tilde{\mu}_{NT,j+1}} = \frac{\mu_{NT,j}}{\mu_{NT,j+1}} + o_p(1) = O_p(1) \text{ for } j = 1,\ldots,R^0-1,$$

where $\mu_{NT,j}$ is the j-th largest eigenvalue of $\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{F'F}{T}\right)$[1] ; and

$$\frac{\tilde{\mu}_{NT,R^0}}{\tilde{\mu}_{NT,R^0+1}} \geq \frac{\mu_{NT,R^0} + O_p\left(N^{-\frac{1}{2}} + m^{-1} + \left(\tilde{\beta}_{(R_{\max})}-\beta\right)\right)}{\left[\bar{c}+o_p(1)\right]/m} \xrightarrow{p} \infty,$$

with $\bar{c} = c_1^2\left(1+\sqrt{y}\right)^2$ and

$$\frac{\tilde{\mu}_{NT,R^0+j}}{\tilde{\mu}_{NT,R^0+j+1}} \leq \frac{\bar{c}+o_p(1)}{\underline{c}+o_p(1)} \text{ for } j = 1,\ldots,\lfloor d^c m\rfloor - 2R^0 - 1,$$

with $\underline{c} = c_2^2 y^{**}\left(1-\sqrt{by^*}\right)^2$ and $y^{**} = \lim_{m\to\infty}\frac{N}{M}$.

As shown in AH, all the reasoning in its proof of Theorem 1 will hold except that $\tilde{\mu}_{NT,j}$ is the j-th largest eigenvalue of $\frac{\widetilde{U}\widetilde{U}'}{NT}$, where $\widetilde{U}$ is the estimator of U. Thus there will be an additional error that measures how precisely U is estimated by $\widetilde{U}$, which is given by $\tilde{\beta}_{(R_{\max})}-\beta$. Following the proof of Lemma A.11 and A.9 in AH, we can see it is sufficient to show that for any $j = 1,\ldots,\lfloor d^c m\rfloor - R^0$,

$$\psi_j\left(\frac{\widetilde{U}\widetilde{U}'}{NT}\right) = \psi_j\left(\frac{UU'}{NT}\right) + o_p(1).$$

Without loss of generality we can focus on the case when the regressor is a scalar, then from

[1] For ease of notation, we drop all the superscripts 0 for true values of factors, factor loadings and regression coefficients in this part of proof.

(3.5.1) we have

$$\frac{1}{\sqrt{NT}}\widetilde{U}=\frac{1}{\sqrt{NT}}X\cdot\left(\beta-\tilde{\beta}_{(R_{\max})}\right)+\frac{1}{\sqrt{NT}}U=\frac{1}{\sqrt{NT}}U+O_p\left(\tilde{\delta}_{R_{\max}}\right),$$

where as before, we denote $\tilde{\delta}_{R_{\max}}=\tilde{\beta}_{(R_{\max})}-\beta$, and $X\cdot\tilde{\delta}_{R_{\max}}=\sum_{p=1}^{P}X_p\tilde{\delta}_{R_{\max},p}$ with X_p and $\tilde{\delta}_{R_{\max},p}$ representing the p-th argument. And

$$\begin{aligned}\frac{\widetilde{U}\widetilde{U}'}{NT}&=\frac{UU'}{NT}-\frac{1}{NT}\left(X\cdot\tilde{\delta}_{R_{\max}}\right)U'-\frac{1}{NT}U\left(X\cdot\tilde{\delta}_{R_{\max}}\right)'+\frac{1}{NT}\left(X\cdot\tilde{\delta}_{R_{\max}}\right)\left(X\cdot\tilde{\delta}_{R_{\max}}\right)'\\&\equiv\frac{UU'}{NT}+R=\frac{UU'}{NT}+O_p\left(\tilde{\delta}_{R_{\max}}\right)\end{aligned}$$

where the order of tail term holds under Frobenius norm and thus folows from Assumption B and C(i). Then Weyl's inequality (or Lemma A.5 in AH) indicates that

$$\psi_j\left(\frac{\widetilde{U}\widetilde{U}'}{NT}\right)\leq\psi_j\left(\frac{UU'}{NT}\right)+\psi_1(R)=\psi_j\left(\frac{UU'}{NT}\right)+O_p\left(\tilde{\delta}_{R_{\max}}\right)\tag{B.1.17}$$

because $|\psi_1(R)|\leq\|R\|$. And next we denote Ξ^k as the matrix of first k-largest eigenvectors of $\frac{UU'}{NT}$ normalized by $\Xi^{k\prime}\Xi^k/T=\mathbb{I}_k$, then for any $k=1,\ldots,R^0$,

$$\begin{aligned}\sum_{j=1}^{k}\psi_j\left(\frac{\widetilde{U}\widetilde{U}'}{NT}\right)&\geq\operatorname{tr}\left(\frac{1}{NT^2}\Xi^{k\prime}UU'\Xi^k+\frac{1}{T}\Xi^{k\prime}R\Xi^k\right)\\&=\sum_{j=1}^{k}\psi_j\left(\frac{UU'}{NT}\right)+O_p\left(\tilde{\delta}_{R_{\max}}\right),\end{aligned}\tag{B.1.18}$$

because $\left|\operatorname{tr}\left(\frac{1}{T}\Xi^{k\prime}R\Xi^k\right)\right|\leq\left\|\frac{1}{\sqrt{T}}\Xi^k\right\|\|R\|$. Note that (B.1.17) and (B.1.18) hold for arbitrary $j,k=1,\ldots,\lfloor d^c m\rfloor-R^0$, which implies that $\psi_j\left(\frac{\widetilde{U}\widetilde{U}'}{NT}\right)=\psi_j\left(\frac{UU'}{NT}\right)+O_p\left(\tilde{\delta}_{R_{\max}}\right)$. And the remaining parts of proof will be the same as in AH using $\frac{UU'}{NT}$, and the consistency of our ER estimator is confirmed by the consistency of $\tilde{\beta}_{(R_{\max})}$, which completes the proof of Theorem 3.5.1. ■

B.2 Proofs of the Technical Lemmas

Proof of Lemma B.1.1. Let $\hat{\delta}=\hat{\beta}-\beta^0$. The proof follows closely that of Proposition A.1(ii) in Bai (2009). By the decomposition in (B.1.2), the fact that $I_9=F\frac{\Lambda'\Lambda}{N}\frac{F^{0\prime}\hat{F}}{T}$, and the definition of H and G, we have

$$\hat{F}V_{NT}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}-F=\hat{F}H^{-1}-F$$

$$= (I_1 + \cdots + I_8) \left(\frac{F^{0\prime}\hat{F}}{T} \right)^{-1} \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} = (I_1 + \cdots + I_8)\, G.$$

Then $T^{-\frac{1}{2}} \left\| \hat{F} V_{NT} \left(\frac{F^{0\prime}\hat{F}}{T} \right)^{-1} \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} - F \right\| = T^{-\frac{1}{2}} \left\| \hat{F} H^{-1} - F \right\| \leq T^{-\frac{1}{2}} (\|I_1\| + \cdots + \|I_8\|) \|G\|$. Note that $T^{-\frac{1}{2}} \left\| \hat{F} \right\| = \sqrt{R}$. As in Bai (2009), it is easy to argue that H is asymptotically nonsingular, so is G. Then $\|G\| = O_p(1)$ and it remains to derive the order of $T^{-\frac{1}{2}} \|I_\ell\|$ for $\ell = 1, ..., 8$. First,

$$\begin{aligned} T^{-\frac{1}{2}} \|I_1\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} X_i \hat{\delta} \hat{\delta}' X_i' \hat{F} \right\| \\ &\leq \frac{1}{N} \sum_{i=1}^{N} \frac{\|X_i\|^2}{T} \left\| \hat{\delta} \right\|^2 T^{-\frac{1}{2}} \left\| \hat{F} \right\| = O_p\left(\left\| \hat{\delta} \right\|^2 \right), \end{aligned}$$

where we use the fact that $\frac{1}{NT} \sum_{i=1}^{N} \|X_i\|^2 = O_p(1)$ by Assumption B(i) and Markov inequality. Next

$$\begin{aligned} T^{-\frac{1}{2}} \|I_2\| = T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} X_i \hat{\delta} \lambda_i' F^{0\prime} \hat{F} \right\| &\lesssim \frac{1}{N} \sum_{i=1}^{N} \frac{\|X_i\|_F \|\lambda_i\|_F}{\sqrt{T}} \left\| \hat{\beta} - \beta \right\| \\ &\lesssim \left(\frac{1}{NT} \sum_{i=1}^{N} \|X_i\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{N} \sum_{i=1}^{N} \|\lambda_i\|^2 \right)^{\frac{1}{2}} \left\| \hat{\delta} \right\| = O_p\left(\left\| \hat{\delta} \right\| \right), \end{aligned}$$

where the last equality holds by Assumption B(iii) and B(iv) and Markov inequality. By the same token, we have $T^{-\frac{1}{2}} \|I_4\| = T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} F^0 \lambda_i \hat{\delta}' X_i' \hat{F} \right\| = O_p\left(\left\| \hat{\delta} \right\| \right)$. For I_3, we have

$$\begin{aligned} T^{-\frac{1}{2}} \|I_3\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} X_i \hat{\delta} \varepsilon_i' \hat{F} \right\| \\ &\leq \frac{1}{N} \sum_{i=1}^{N} \frac{\|X_i\| \|\varepsilon_i\|}{T} \left\| \hat{\delta} \right\| T^{-\frac{1}{2}} \left\| \hat{F} \right\| \\ &\lesssim \left(\frac{1}{NT} \sum_{i=1}^{N} \|X_i\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{NT} \sum_{i=1}^{N} \|\varepsilon_i\|^2 \right)^{\frac{1}{2}} \left\| \hat{\delta} \right\| = O_p\left(\left\| \hat{\delta} \right\| \right), \end{aligned}$$

and similarly, $T^{-\frac{1}{2}} \|I_5\| = T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \hat{\delta}' X_i' \hat{F} \right\| = O_p\left(\left\| \hat{\delta} \right\| \right)$. For I_6, we have

$$T^{-1} \|I_6\|^2 = \frac{1}{T} \left\| \frac{1}{NT} F \sum_{i=1}^{N} \lambda_i \varepsilon_i' \hat{F} \right\|^2 \lesssim \frac{1}{N} \left(\frac{1}{NT} \left\| \sum_{i=1}^{N} \lambda_i \varepsilon_i' \right\|^2 \right) = O_p\left(N^{-1} \right),$$

where the last equality holds by the fact that

$$E\left(\frac{1}{NT} \left\| \sum_{i=1}^{N} \lambda_i \varepsilon_i' \right\|^2 \right) = \frac{1}{NT} \sum_{i,j=1}^{N} E\left(\varepsilon_i' \varepsilon_j \right) E(\lambda_j' \lambda_i) \leq \frac{M}{N} \sum_{i,j=1}^{N} \overline{\sigma}_{ij} = O(1)$$

by Assumption A(v), B(iv) and C(ii). Therefore $T^{-\frac{1}{2}}\|I_6\| = O_p(N^{-\frac{1}{2}})$. Analogously, $T^{-1}\|I_7\|^2 = \frac{1}{T}\left\|\frac{1}{NT}\sum_{i=1}^N \varepsilon_i\lambda_i' F^{0\prime}\hat{F}\right\|^2 = O_p\left(N^{-1}\right)$.Note that so far the orders for terms I_1-I_7 all replicate those in Bai (2009, pp. 1267).

Now, we study I_8. Let I_{8t} denote the t-th row of I_8, which can be decomposed as follows:

$$
\begin{aligned}
I_{8t} &= \frac{1}{NT}\sum_{i=1}^N \varepsilon_{it}\varepsilon_i'\hat{F} = \frac{1}{T}\sum_{s=1}^T \frac{1}{N}\sum_{i=1}^N \varepsilon_{it}\varepsilon_{is}\hat{F}_s' \\
&= \frac{1}{T}\sum_{s=1}^T \gamma_N(s,t)\hat{F}_s' + \frac{1}{T}\sum_{s=1}^T\left(\frac{1}{N}\sum_{i=1}^N \varepsilon_{it}\varepsilon_{is} - \gamma_N(s,t)\right)\hat{F}_s' \equiv I_{8t,1} + I_{8t,2},
\end{aligned}
$$

where $\gamma_N(s,t) = \frac{1}{N}\sum_{i=1}^N E(\varepsilon_{it}\varepsilon_{is})$. Then $T^{-1}\|I_8\|^2 = T^{-1}\sum_{t=1}^T\|I_{8t}\|^2 \le 2T^{-1}\sum_{t=1}^T\|I_{8t,1}\|^2 + 2T^{-1}\sum_{t=1}^T\|I_{8t,2}\|^2 = II_1 + II_2$. By Cauchy-Schwarz inequality and by Assumption C(ii),

$$
II_1 \le \frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \gamma_N(s,t)^2\right)\left(\frac{1}{T}\sum_{s=1}^T\left\|\hat{F}_s'\right\|_F^2\right) \lesssim \frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \gamma_N(s,t)^2 = O_p\left(T^{\max(4d_\varepsilon,1)-2}\right),
$$

For II_2, we have

$$
\begin{aligned}
II_2 &= \frac{1}{T}\sum_{t=1}^T\left\|\frac{1}{T}\sum_{s=1}^T \xi_{st}\hat{F}_s'\right\|^2 = \frac{1}{T}\frac{1}{T^2}\sum_{s,u=1}^T \hat{F}_s'\hat{F}_u\sum_{t=1}^T \xi_{st}\xi_{ut} \\
&\le \frac{1}{T}\left(\frac{1}{T^2}\sum_{s,u=1}^T\left(\hat{F}_s'\hat{F}_u\right)^2\right)^{\frac{1}{2}}\left(\frac{1}{T^2}\sum_{s,u=1}^T\left(\sum_{t=1}^T \xi_{st}\xi_{ut}\right)^2\right)^{\frac{1}{2}} = O_p(N^{-1}),
\end{aligned}
$$

where $\xi_{st} = \frac{1}{N}\sum_{i=1}^N \varepsilon_{it}\varepsilon_{is} - \gamma_N(s,t)$, and the last equality above holds by Assumption B(iii) and the fact that

$$
\begin{aligned}
\frac{1}{T^2}\sum_{s,u=1}^T E\left(\sum_{t=1}^T \xi_{st}\xi_{ut}\right)^2 &\le \frac{1}{T^2}\sum_{s,u=1}^T T^2\max_{t,v} E|\xi_{vt}|^4 \qquad \text{(B.2.1)} \\
&= \frac{T^2}{N^2}\max_{t,v} E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\varepsilon_{it}\varepsilon_{iv} - E(\varepsilon_{it}\varepsilon_{iv})\right]\right|^4 = O\left(\frac{T^2}{N^2}\right)
\end{aligned}
$$

by Assumption C(iii). Then $T^{-1}\|I_8\|^2 = O_p\left(N^{-1} + T^{\max(4d_\varepsilon,1)-2}\right)$. Therefore by the invertibility of H, we can conclude that $\frac{1}{T}\left\|\hat{F} - F^0H\right\|^2 = O_p\left(\left\|\hat{\delta}\right\|^2\right) + O_p\left(\delta_{NT}^{-2}\right)$. ■

Proof of Lemma B.1.2. Note that $\frac{1}{N}\sum_{i=1}^N \lambda_i\frac{\varepsilon_i'\hat{F}}{T} = \frac{1}{N}\sum_{i=1}^N \lambda_i\frac{\varepsilon_i'F^0H}{T} + \frac{1}{N}\sum_{i=1}^N \lambda_i\frac{\varepsilon_i'(\hat{F}-F^0H)}{T} \equiv A_1 + A_2$. For A_1, we have

$$
\left\|\frac{1}{N}\sum_{i=1}^N \lambda_i\frac{\varepsilon_i'F^0H}{T}\right\| \lesssim \left\|\frac{1}{N}\sum_{i=1}^N \lambda_i\frac{\varepsilon_i'F^0}{T}\right\| \equiv \left\|\bar{A}_1\right\|.
$$

Note that

$$
\begin{aligned}
E\left\|\bar{A}_1\right\|^2 &= \frac{1}{N^2T^2}\sum_{i,j=1}^{N} E\left[\left(\varepsilon_i' F^0 F^{0\prime}\varepsilon_j\right)\left(\lambda_j'\lambda_i\right)\right] = \frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} E\left(\varepsilon_{it}\varepsilon_{js}\right) E\left(F_t^{0\prime} F_s \lambda_j'\lambda_i\right) \\
&\leq \frac{\max_{i,j,t,s}\left|E\left(F_t^{0\prime} F_s\lambda_j'\lambda_i\right)\right|}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{js}\right)\right| \\
&\leq \frac{M}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{is}\right)\right| = O\left(N^{-1}T^{2d_\varepsilon-1}\right), \qquad \text{(B.2.2)}
\end{aligned}
$$

where the second quality holds by Assumption A(v), and the second inequality can be derived from Assumption B(iii) and B(iv) using Cauchy-Schwarz inequality, and the last equality holds by Assumption C(ii). Then $A_1 = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\right)$. For A_2, following the proof of Lemma B.1.1 and recalling that $G = \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}$

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{\varepsilon_i'\left(\hat{F}-F^0H\right)}{T} &= \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{\varepsilon_i'\left(\hat{F}H^{-1}-F\right)}{T}H = \frac{1}{NT}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\left(I_1+\cdots+I_8\right)GH \\
&\equiv \left(a_1+\cdots+a_8\right)GH.
\end{aligned}
$$

Then it remains to bound a_ℓ's by following partly the proof of Lemma A.4(ii) in Bai (2009) and use some results derived in our proof of Lemma B.1.1. For a_1 we have

$$
\begin{aligned}
\left\|a_1\right\| &= \left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\varepsilon_i' X_k\hat{\delta}\hat{\delta}' X_k'\hat{F}G\right\| \\
&\leq T^{-\frac{1}{2}}\left\|\hat{\delta}\right\|^2\frac{1}{NT}\sum_{k=1}^{N}\left(\frac{1}{N}\left\|\sum_{i=1}^{N}\lambda_i\varepsilon_i' X_k\right\|\left\|X_k\right\|\right)T^{-\frac{1}{2}}\left\|\hat{F}\right\|\left\|G\right\| \\
&\lesssim T^{-\frac{1}{2}}\left\|\hat{\delta}\right\|^2\left(\frac{1}{NT}\sum_{k=1}^{N}\frac{1}{N^2}\left\|\sum_{i=1}^{N}\lambda_i\varepsilon_i' X_k\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|X_k\right\|^2\right)^{\frac{1}{2}} \\
&\lesssim T^{-\frac{1}{2}}\left\|\hat{\delta}\right\|^2\left(\frac{1}{NT}\sum_{k=1}^{N}\frac{1}{N^2}\left\|\sum_{i=1}^{N}\lambda_i\varepsilon_i' X_k\right\|^2\right)^{\frac{1}{2}} = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left\|\hat{\delta}\right\|^2\right)
\end{aligned}
$$

where we use the result that

$$
\begin{aligned}
E\left(\frac{1}{NT}\sum_{k=1}^{N}\frac{1}{N^2}\left\|\sum_{i=1}^{N}\lambda_i\varepsilon_i' X_k\right\|^2\right) &= \frac{1}{NT}\sum_{k=1}^{N}\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}E\left(\varepsilon_{it}\varepsilon_{js}\right)E\left(\lambda_i'\lambda_j X_{kt}' X_{ks}\right) \\
&\leq \max_{i,j,t,s}\frac{1}{N}\sum_{k=1}^{N}\left|E\left(\lambda_i'\lambda_j X_{kt}' X_{ks}\right)\right|\frac{1}{N^2T}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{js}\right)\right| \\
&\leq \frac{M}{N^2T}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{js}\right)\right| = O\left(N^{-1}T^{2d_\varepsilon}\right)
\end{aligned}
$$

by Assumption B(i), B(iv) and C(ii). Similarly,

$$\begin{aligned}
\|a_2\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \left(\frac{1}{NT} \sum_{k=1}^{N} X_k \hat{\delta} \lambda_k' F^{0\prime} \hat{F} \right) G \right\| \\
&= \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{1}{N} \sum_{k=1}^{N} \lambda_i \varepsilon_i' X_k \hat{\delta} \lambda_k' \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \right\| \\
&\lesssim \frac{1}{NT} \sum_{k=1}^{N} \left(\frac{1}{N} \left\| \sum_{i=1}^{N} \lambda_i \varepsilon_i' X_k \right\| \|\lambda_k\| \right) \left\| \hat{\delta} \right\| \\
&\leq T^{-\frac{1}{2}} \left\| \hat{\delta} \right\| \left(\frac{1}{NT} \sum_{k=1}^{N} \frac{1}{N^2} \left\| \sum_{i=1}^{N} \lambda_i \varepsilon_i' X_k \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{N} \sum_{k=1}^{N} \|\lambda_k\|^2 \right)^{\frac{1}{2}} \\
&= O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left\| \hat{\delta} \right\| \right),
\end{aligned}$$

and

$$\begin{aligned}
\|a_3\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \left(\frac{1}{NT} \sum_{k=1}^{N} X_k \hat{\delta} \varepsilon_k' \hat{F} \right) G \right\| \\
&\leq T^{-\frac{1}{2}} \frac{1}{NT} \sum_{k=1}^{N} \left(\left\| \frac{1}{N} \sum_{i=1}^{N} \lambda_i \varepsilon_i' X_k \right\| \|\varepsilon_k\| \right) \left\| \hat{\delta} \right\| T^{-\frac{1}{2}} \left\| \hat{F} \right\| \|G\| \\
&\lesssim T^{-\frac{1}{2}} \left\| \hat{\delta} \right\| \left(\frac{1}{NT} \sum_{k=1}^{N} \frac{1}{N^2} \left\| \sum_{i=1}^{N} \lambda_i \varepsilon_i' X_k \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{NT} \sum_{k=1}^{N} \|\varepsilon_k\|^2 \right)^{\frac{1}{2}} \\
&= O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left\| \hat{\delta} \right\| \right).
\end{aligned}$$

Next,

$$\begin{aligned}
\|a_4\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \left(\frac{1}{NT} \sum_{k=1}^{N} F \lambda_k \hat{\delta}' X_k' \hat{F} \right) G \right\| \\
&\leq T^{-\frac{1}{2}} \frac{1}{NT} \sum_{k=1}^{N} \left(\left\| \frac{1}{N} \sum_{i=1}^{N} \lambda_i \varepsilon_i' F \right\| \|\lambda_k\| \|X_k\| \right) \left\| \hat{\delta} \right\| T^{-\frac{1}{2}} \left\| \hat{F} \right\| \|G\| \\
&= O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left\| \hat{\delta} \right\| \right)
\end{aligned}$$

where the last equality holds by using Cauchy-Schwarz inequality, Assumption B(i) and B(iv), and the same reasoning to obtain the order of $\bar{A}_1$ above. For a_5 we have

$$\begin{aligned}
\|a_5\| &= \left\| \frac{1}{NT} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \varepsilon_i' \varepsilon_k \hat{\delta}' \frac{X_k' \hat{F}}{T} G \right\| \\
&= \left\| \frac{1}{NT} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right) \left(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \varepsilon_{kt} \hat{\delta}' \frac{X_k' \hat{F}}{T} \right) G \right\| \\
&\leq \frac{1}{NT} \sum_{t=1}^{T} \left(\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right\| \right) \left(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \|\varepsilon_{kt}\| \frac{\|X_k\|}{\sqrt{T}} \right) \left\| \hat{\delta} \right\| T^{-\frac{1}{2}} \left\| \hat{F} \right\| \|G\|
\end{aligned}$$

$$\lesssim \left\|\hat{\delta}\right\| \left(\frac{1}{NT}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{it}\right\|^2\right)^{\frac{1}{2}} \left(\frac{1}{NT}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\|\varepsilon_{kt}\|\frac{\|X_k\|}{\sqrt{T}}\right)^2\right)^{\frac{1}{2}}$$
$$\equiv \left\|\hat{\delta}\right\|\sqrt{a_{51}}\sqrt{a_{52}}.$$

Note that

$$E\left(a_{51}\right)=\frac{1}{NT}\sum_{t=1}^{T}\frac{1}{N}\sum_{i,j=1}^{N}E\left(\lambda_i'\lambda_j\right)E\left(\varepsilon_{it}\varepsilon_{jt}\right)\leq \max_{i,j}\left|E\left(\lambda_i'\lambda_j\right)\right|\frac{1}{NT}\sum_{t=1}^{T}\frac{1}{N}\sum_{i,j=1}^{N}\left|E\left(\varepsilon_{it}\varepsilon_{jt}\right)\right|=O\left(N^{-1}\right)$$

by Assumption B(iv) and C(ii), and

$$a_{52}\leq \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{k=1}^{N}\|\varepsilon_{kt}\|^2\right)\left(\frac{1}{N}\sum_{k=1}^{N}\frac{\|X_k\|^2}{T}\right)=O_p\left(1\right)$$

by Cauchy-Schwarz inequality, Assumption B(i) and C(i). Then $\|a_5\|=O_p\left(N^{-1/2}\left\|\hat{\delta}\right\|\right)$. For a_6, we have

$$a_6=\frac{1}{NT}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\left(\frac{1}{NT}\sum_{k=1}^{N}F\lambda_k\varepsilon_k'\hat{F}\right)G$$
$$=\frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\sum_{k=1}^{N}F\lambda_k\varepsilon_k'\left[F^0H+\left(\hat{F}-F^0H\right)\right]G\equiv a_{61}+a_{62}.$$

Note that $E\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i\varepsilon_i'F\right\|^2=\frac{1}{N^2T^2}\sum_{i,j=1}^{N}E(\varepsilon_{it}\varepsilon_{is})E(\lambda_i'\lambda_jF_t'F_s)=O_p\left(N^{-1}T^{2d_\varepsilon-1}\right)$ by using the same reasoning as we analyze $\bar{A}_1$ above, we have

$$\|a_{61}\|=\left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_i\varepsilon_i'F\sum_{k=1}^{N}\lambda_k\varepsilon_k'FHG\right\|\lesssim\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i\varepsilon_i'F\right\|^2=O_p\left(N^{-1}T^{2d_\varepsilon-1}\right),$$

and

$$\|a_{62}\|=\left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_i\varepsilon_i'F\sum_{k=1}^{N}\lambda_k\varepsilon_k'\left(\hat{F}-F^0H\right)G\right\|$$
$$\lesssim\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i\varepsilon_i'F\right\|\left\|\frac{1}{NT}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\left(\hat{F}-F^0H\right)\right\|$$
$$\leq O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\lambda_k\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{k=1}^{N}\|\varepsilon_k\|^2\right)^{\frac{1}{2}}\frac{\left\|\hat{F}-F^0H\right\|}{\sqrt{T}}$$
$$=O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$$

by Lemma B.1.1. Then $\|a_6\|=O_p\left(N^{-1}T^{2d_\varepsilon-1}+N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$. Next, for a_7

we have

$$
\begin{aligned}
\|a_7\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \left(\frac{1}{NT} \sum_{k=1}^{N} \varepsilon_k \lambda_k' F^{0\prime} \hat{F} \right) G \right\| \\
&= \left\| \frac{1}{N^2 T} \sum_{k=1}^{N} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \varepsilon_k \lambda_k' \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \right\| \\
&\lesssim \frac{1}{NT} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right\|_F \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \varepsilon_{kt} \lambda_k' \right\|_F \\
&\leq \frac{1}{NT} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right\|^2 = O_p\left(N^{-1}\right)
\end{aligned}
$$

using the same reasoning as above by Assumption B(iv) and C(ii).

Lastly, we study a_8 by making the following decomposition

$$
a_8 = \frac{1}{N^2 T^2} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \sum_{k=1}^{N} \varepsilon_k \varepsilon_k' \left[F^0 H + \left(\hat{F} - F^0 H \right) \right] G \equiv a_{81} + a_{82}.
$$

For a_{81}, we have

$$
\begin{aligned}
a_{81}(HG)^{-1} &= \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \sum_{t=1}^{T} \varepsilon_{it} \varepsilon_{kt} \sum_{s=1}^{T} \varepsilon_{ks} F_s \\
&= \frac{1}{T\sqrt{N}} \frac{1}{N} \sum_{k=1}^{N} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_i [\varepsilon_{it} \varepsilon_{kt} - E\left(\varepsilon_{it} \varepsilon_{kt}\right)] \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{ks} F_s \right) \\
&\quad + \frac{1}{N\sqrt{T}} \frac{1}{N} \sum_{k=1}^{N} \left(\frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_i E\left(\varepsilon_{it} \varepsilon_{kt}\right) \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{ks} F_s \right) \equiv a_{81a} + a_{81b}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\|a_{81a}\| &\leq \frac{1}{T\sqrt{N}} \frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_i [\varepsilon_{it} \varepsilon_{kt} - E\left(\varepsilon_{it} \varepsilon_{kt}\right)] \right\| \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{ks} F_s \right\| \\
&\leq O_p\left(T^{d_\varepsilon}\right) \frac{1}{T\sqrt{N}} \left(\frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_i [\varepsilon_{it} \varepsilon_{kt} - E\left(\varepsilon_{it} \varepsilon_{kt}\right)] \right\|^2 \right)^{\frac{1}{2}}
\end{aligned}
$$

by arguments as used in the analysis of A_1 above and Assumption C(ii). By Assumption C(iv),

$$
\begin{aligned}
& E\left(\frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_i [\varepsilon_{it} \varepsilon_{kt} - E\left(\varepsilon_{it} \varepsilon_{kt}\right)] \right\|^2 \right) \\
&= \frac{1}{N} \sum_{k=1}^{N} \frac{1}{NT} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} E\left(\lambda_i' \lambda_j\right) E\left\{ [\varepsilon_{it} \varepsilon_{kt} - E\left(\varepsilon_{it} \varepsilon_{kt}\right)][\varepsilon_{js} \varepsilon_{ks} - E\left(\varepsilon_{js} \varepsilon_{ks}\right)] \right\} \\
&\leq \frac{M}{N^2 T} \sum_{i,j,k=1}^{N} \sum_{t,s=1}^{T} \left| \operatorname{cov}\left(\varepsilon_{it} \varepsilon_{kt}, \varepsilon_{js} \varepsilon_{ks}\right) \right| = O\left(T^{2d_\varepsilon}\right).
\end{aligned}
$$

It follows that that $\|a_{81a}\| = O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon-1}\right)$. Next, noting that

$$\|a_{81b}\| \le \frac{1}{N\sqrt{T}}\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{N}\|\lambda_i\|\left(\frac{1}{T}\sum_{t=1}^{T}|E(\varepsilon_{it}\varepsilon_{kt})|\right)\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{ks}F_s\right\|,$$

$$\begin{aligned}
E\|a_{81b}\| &\le \left(\max_k E\left(\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{ks}F_s\right\|^2\right)^{\frac{1}{2}}\right)\frac{1}{N^2\sqrt{T}}\sum_{k=1}^{N}\sum_{i=1}^{N}E\left(\|\lambda_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{t=1}^{T}E\left|(\varepsilon_{it}\varepsilon_{kt})\right|\right)\\
&\le MT^{d_\varepsilon}\frac{1}{N^2\sqrt{T}}\sum_{k=1}^{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}|E(\varepsilon_{it}\varepsilon_{kt})| = O\left(N^{-1}T^{d_\varepsilon-\frac{1}{2}}\right)
\end{aligned}$$

by Cauchy-Schwarz inequality, the reasoning for A_1 above, and Assumption C(ii). Then $\|a_{81b}\| = O_p\left(N^{-1}T^{d_\varepsilon-\frac{1}{2}}\right)$ and $\|a_{81}\| = O_p(N^{-\frac{1}{2}}T^{2d_\varepsilon-1} + N^{-1}T^{d_\varepsilon-\frac{1}{2}})$. Next we analyze a_{82}:

$$\begin{aligned}
a_{82}G^{-1} &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\varepsilon_i'\varepsilon_k\varepsilon_k'\left(\hat{F}-F^0H\right)\\
&= \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{k=1}^{N}\xi_k\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T} + \frac{1}{N}\frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_iE(\varepsilon_{it}\varepsilon_{kt})\right)\left(\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right)\\
&\equiv a_{82a} + a_{82b}
\end{aligned}$$

where $\xi_k = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i[\varepsilon_{it}\varepsilon_{kt} - E(\varepsilon_{it}\varepsilon_{kt})]$. It is easy to show that $\frac{1}{N}\sum_{k=1}^{N}E\|\xi_k\|^2 = O\left(T^{2d_\varepsilon}\right)$ under Assumptions A(v), B(iv) and C(iv). Then by Lemma B.1.1,

$$\begin{aligned}
\|a_{82a}\| &= \frac{1}{\sqrt{NT}}\left\|\frac{1}{N}\sum_{k=1}^{N}\xi_k\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right\|\\
&\le \frac{1}{\sqrt{NT}}\left\{\frac{1}{N}\sum_{k=1}^{N}\|\xi_k\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right\|^2\right\}^{1/2}\\
&= (NT)^{-1/2}O_p\left(T^{d_\varepsilon}\right)O_p\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right),
\end{aligned}$$

and

$$\begin{aligned}
\|a_{82b}\| &\le \frac{1}{N}\left\|\frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_iE(\varepsilon_{it}\varepsilon_{kt})\right)\left(\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right)\right\|\\
&\le \frac{1}{N}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_iE(\varepsilon_{it}\varepsilon_{kt})\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right\|^2\right\}^{1/2}\\
&= N^{-1}O_p(1)\,O_p\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right).
\end{aligned}$$

So $\|a_{82}\| = O_p\left((N^{-\frac{1}{2}}T^{d_\varepsilon-1/2}+N^{-1})\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$ and $\|a_8\| = O_p(N^{-\frac{1}{2}}T^{2d_\varepsilon-1}+N^{-1}T^{d_\varepsilon-\frac{1}{2}})+O_p\left((N^{-\frac{1}{2}}T^{d_\varepsilon-1/2}+N^{-1})\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$.

In sum, we can conclude that

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{\varepsilon_i'\hat{F}}{T}\right\| = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}+N^{-1}+N^{-\frac{1}{2}}\left\|\hat{\delta}\right\|\right),$$

which then completes the proof of Lemma B.1.2. ■

Proof of Lemma B.1.3. Consider the following decomposition of $J8$:

$$\begin{aligned}J_8 &= \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\Omega_k\hat{F}G\lambda_i+\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)\hat{F}G\lambda_i\\ &\equiv J_{81}+J_{82},\end{aligned}$$

where $\Omega_k = E\left(\varepsilon_k\varepsilon_k'\right)$ and $J_{81} = A_{NT}$. For J_{81}, we have

$$\begin{aligned}\|J_{81}\| &= \left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\Omega_k\hat{F}G\lambda_i\right\| \lesssim \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|X_i\|\,\|\lambda_i\|\left\|\frac{1}{NT}\sum_{k=1}^{N}\Omega_k\right\|\\ &\lesssim \left(\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\sum_{t,s=1}^{T}E\left(\varepsilon_{it}\varepsilon_{is}\right)E\left(\varepsilon_{kt}\varepsilon_{ks}\right)\right)^{\frac{1}{2}}\\ &\leq \left(\frac{1}{T^2}\sum_{t,s=1}^{T}|\gamma_N(s,t)|^2\right)^{\frac{1}{2}} = O_p\left(T^{\max(2d_\varepsilon,1/2)-1}\right),\end{aligned}$$

by Assumption B(i), B(iv) and C(ii). For J_{82}, we make the decomposition

$$\begin{aligned}J_{82} &= \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)FHG\lambda_i\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)(\hat{F}-F^0H)G\lambda_i \equiv J_{821}+J_{822}.\end{aligned}$$

For J_{821}, we have

$$\begin{aligned}J_{821} &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left[X_i'\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)FHG\lambda_i\right]-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left[X_i'\frac{1}{T}\hat{F}\hat{F}'\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)FHG\lambda_i\right]\\ &= \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\zeta_iFHG\lambda_i-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left[X_i'\frac{1}{T}\hat{F}\hat{F}'\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)FHG\lambda_i\right]\\ &\equiv J_{821a}+J_{821b},\end{aligned}$$

where

$$\zeta_i = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}X_{it}\left(\varepsilon_{kt}\varepsilon_{ks}-E\left(\varepsilon_{kt}\varepsilon_{ks}\right)\right)F_s'.$$

Note that

$$E\left\|\zeta_i\right\|^2 = \frac{1}{N}\sum_{i,k=1}^{N}\frac{1}{T^2}\sum_{t,s,u,v=1}^{T}E\left(X_{it}'X_{iu}F_s'F_v\right)\operatorname{cov}\left(\varepsilon_{kt}\varepsilon_{ks},\varepsilon_{iu}\varepsilon_{iv}\right)$$

$$\leq \max_{i,k,t,s,u,v}\left|E\left(X_{it}'X_{iu}F_s'F_v\right)\right|\frac{1}{N}\sum_{i,k=1}^{N}\frac{1}{T^2}\sum_{t,s,u,v=1}^{T}\left|\operatorname{cov}\left(\varepsilon_{kt}\varepsilon_{ks},\varepsilon_{iu}\varepsilon_{iv}\right)\right| = O\left(T^{4d_\varepsilon}\right),$$

by Assumption C(iv). With this, we can readily show that $J_{821a} = O_p(N^{-\frac{1}{2}}T^{2d_\varepsilon-1})$. For J_{821b}, we have

$$\begin{aligned}\left\|J_{821b}\right\| &\lesssim \frac{1}{\sqrt{NT}}\frac{1}{NT}\sum_{i=1}^{N}\left\|X_i'\hat{F}\right\|\left\|\lambda_i\right\|\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_t\left[\varepsilon_{kt}\varepsilon_{ks}-E\left(\varepsilon_{kt}\varepsilon_{ks}\right)\right]F_s'\right\| \\ &\lesssim \frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_t\left[\varepsilon_{kt}\varepsilon_{ks}-E\left(\varepsilon_{kt}\varepsilon_{ks}\right)\right]F_s'\right\| \\ &\lesssim \frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t\left[\varepsilon_{kt}\varepsilon_{ks}-E\left(\varepsilon_{kt}\varepsilon_{ks}\right)\right]F_s'\right\| \\ &+ \frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_t-HF_t\right)\left[\varepsilon_{kt}\varepsilon_{ks}-E\left(\varepsilon_{kt}\varepsilon_{ks}\right)\right]F_s'\right\| \\ &\equiv \frac{1}{\sqrt{NT}}\left\{J_{821b1}+J_{821b2}\right\}.\end{aligned}$$

Using the same reasoning as used for J_{821a}, we can show $J_{821b1} = O_p(N^{-\frac{1}{2}}T^{2d_\varepsilon-1})$. In addition, by Lemma B.1.1 and the fact $\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)\right\| = O_p(T^{\frac{1}{2}+d_\varepsilon})$ under Assumption C,

$$\left\|J_{821b2}\right\| \lesssim \frac{1}{\sqrt{T}}\left\|\hat{F}-F^0H\right\|\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)\right\| = O_p\left(T^{\frac{1}{2}+d_\varepsilon}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right).$$

Then $J_{821b} = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$. Next, for J_{822} we have

$$\begin{aligned}\left\|J_{822}\right\| &= \left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)\left(\hat{F}-F^0H\right)G\lambda_i\right\| \\ &\lesssim \left\|\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)\right\|\frac{1}{\sqrt{T}}\left\|\hat{F}-F^0H\right\| = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right).\end{aligned}$$

In sum, we have

$$J_8 = A_{NT} + O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon-1} + N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right),$$

which finishes the proof of Lemma B.1.3. ■

Proof of Lemma B.1.4. Following the proof of Lemma A.8 in Bai (2009), we first study

$$\frac{1}{NT}\sum_{i=1}^{N} X_i' \left(\mathbf{M}_F - \mathbf{M}_{\hat{F}}\right)\varepsilon_i.$$

We make the following decomposition

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N} X_i' \left(\mathbf{M}_F - \mathbf{M}_{\hat{F}}\right)\varepsilon_i \\
&= \frac{1}{NT}\sum_{i=1}^{N} X_i' \left(\mathbf{P}_{\hat{F}} - \mathbf{P}_F\right)\varepsilon_i \\
&= \frac{1}{NT}\sum_{i=1}^{N} \frac{X_i'\left(\hat{F}-F^0H\right)}{T} H'F'\varepsilon_i + \frac{1}{NT}\sum_{i=1}^{N} \frac{X_i'\left(\hat{F}-F^0H\right)}{T}\left(\hat{F}-F^0H\right)'\varepsilon_i \\
&+ \frac{1}{NT}\sum_{i=1}^{N} \frac{X_i'F^0H}{T}\left(\hat{F}-F^0H\right)'\varepsilon_i + \frac{1}{NT}\sum_{i=1}^{N} \frac{X_i'F}{T}\left[HH' - \left(\frac{F'F}{T}\right)^{-1}\right]F'\varepsilon_i \\
&\equiv a+b+c+d.
\end{aligned}$$

For a, we have

$$\begin{aligned}
\|a\| &= \left\| \frac{1}{T}\sum_{s=1}^{T}\left(\hat{F}_s - H'F_s\right)' H' \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t X_{is}\varepsilon_{it}\right)\right\| \\
&\lesssim \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - H'F_s\right\|^2\right)^{\frac{1}{2}} \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t X_{is}\varepsilon_{it}\right\|^2\right)^{\frac{1}{2}} \\
&= O_p\left(N^{-1/2}T^{d_\varepsilon - 1/2}\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right)\right)
\end{aligned}$$

by Lemma B.1.1 and the fact that

$$\begin{aligned}
&E\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t X_{is}\varepsilon_{it}\right\|^2\right) \\
&= \frac{1}{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{r,t=1}^{T} E\left(\varepsilon_{it}\varepsilon_{jr}\right)E\left(F_t'F_rX_{is}'X_{js}\right) \\
&\le \max_{i,j,t,r,s} E\left(F_t'F_rX_{is}'X_{js}\right)\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{r,t=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{jr}\right)\right| = O\left(N^{-1}T^{2d_\varepsilon - 1}\right)
\end{aligned}$$

under Assumptions B(i), B(iv) and C(ii). Next, for b we have

$$\begin{aligned}
\|b\| &= \left\|\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'\left(\hat{F}-F^0H\right)}{T}\left(\hat{F}-F^0H\right)'\varepsilon_i\right\|_F \\
&\le \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - H'F_t\right\|_F^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}X_{is}\varepsilon_{it}\right\|^2\right)^{\frac{1}{2}}
\end{aligned}$$

$$= O_p\left(N^{-1/2}\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right)\right)$$

by Cauchy-Schwarz inequality, Lemma B.1.1 and the fact that

$$E\left(\frac{1}{T^2}\sum_{t,s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}X_{is}\varepsilon_{it}\right\|^2\right) = \frac{1}{T^2}\sum_{s=1}^{T}\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t=1}^{T}E\left(\varepsilon_{it}\varepsilon_{jt}\right)E\left(X_{is}'X_{js}\right)$$
$$\leq \max_{i,j,s}\left|E\left(X_{is}'X_{js}\right)\right|\frac{1}{N^2T}\sum_{i,j=1}^{N}\sum_{t=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{jt}\right)\right| = O\left(N^{-1}\right).$$

Next, we study c by making the following decomposition:

$$c = \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\left(\frac{F'F}{T}\right)^{-1}\left(\hat{F}H^{-1} - F\right)'\varepsilon_i$$
$$+ \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\left(HH' - \left(\frac{F'F}{T}\right)^{-1}\right)\left(\hat{F}H^{-1} - F\right)'\varepsilon_i \equiv c_1 + c_2.$$

For c_2 we have, by denoting $Q = HH' - \left(\frac{F'F}{T}\right)^{-1}$ that

$$\|c_2\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}Q\left(\hat{F}H^{-1} - F\right)'\varepsilon_i\right\|$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\left[\varepsilon_i'\left(\hat{F}H^{-1} - F\right)\otimes\left(\frac{X_i'F}{T}\right)\right]\text{vec}\left(Q\right)$$
$$= \left[O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon - \frac{1}{2}}\right) + O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon - \frac{1}{2}}\left\|\hat{\delta}\right\|\right)\right]\text{vec}\left(Q\right)$$

by the proof of Lemma B.1.2. Next by Assumption B(iii) and Lemma B.1.1,

$$\frac{1}{T}F'\left(\hat{F} - F^0H\right) = O_p\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right),$$

and the same order holds for $\frac{1}{T}\hat{F}'\left(\hat{F} - F^0H\right)$. Then left multiplying $\frac{1}{T}F'\left(\hat{F} - F^0H\right)$ by H' and using the transpose of $\frac{1}{T}\hat{F}'\left(\hat{F} - F^0H\right)$, we can obtain

$$\mathbb{I}_R - H'\frac{F'F}{T}H = O_p\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right),$$

where the same order holds for $\mathbb{I}_R - \frac{F'F}{T}HH'$ and thus for Q. Therefore $c_2 = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon - \frac{1}{2}}\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right)\right)$.

For c_1, we have by (B.1.2) and the proof of Lemma B.1.1 that

$$c_1 = \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\left(\frac{F'F}{T}\right)^{-1}\left(\hat{F}H^{-1} - F\right)'\varepsilon_i$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\left(\frac{F'F}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\left(\frac{\hat{F}'F}{T}\right)^{-1}\left(I_1 + \cdots + I_8\right)'\varepsilon_i \equiv c_{1,1} + \cdots + c_{1,8}.$$

For c_{11} we have, by denoting $\tilde{G} = \left(\frac{F'F}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\left(\frac{\hat{F}'F}{T}\right)^{-1}$ that

$$\begin{aligned}
\|c_{1,1}\| &= \left\| \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}I_1'\varepsilon_i \right\| \\
&= \left\| \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\hat{F}'X_k\hat{\delta}\hat{\delta}'X_k'\varepsilon_i \right\| \\
&\lesssim \left\|\hat{\delta}\right\|^2 \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\left\|\frac{X_i'F}{T}\right\|\|\varepsilon_i\|\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\frac{1}{NT}\sum_{k=1}^{N}\|X_k\|^2 = O_p\left(\left\|\hat{\delta}\right\|^2\right)
\end{aligned}$$

by Assumption B(i), B(iii) and C(ii), and the fact that $\tilde{G} = O_p(1)$. For $c_{1,2}$, we have

$$\begin{aligned}
\|c_{1,2}\| &= \left\| \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\hat{F}'F\lambda_k\hat{\delta}'X_k'\varepsilon_i \right\| \\
&\lesssim \left\|\hat{\delta}\right\| \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{X_i'F}{T}\right\|\frac{1}{NT}\left\|\sum_{k=1}^{N}X_k'\varepsilon_i\lambda_k'\right\| \\
&\leq \left\|\hat{\delta}\right\|\frac{1}{\sqrt{N}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{X_i'F}{T}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N^2T^2}\sum_{i=1}^{N}\left\|\sum_{k=1}^{N}X_k'\varepsilon_i\lambda_k'\right\|^2\right)^{\frac{1}{2}} \\
&= O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-1/2}\left\|\hat{\delta}\right\|\right),
\end{aligned}$$

by Cauchy-Schwarz inequality and similar arguments as used above. Similarly, $c_{1,\ell} = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left\|\hat{\delta}\right\|\right)$ for $\ell = 3,4,5$ as in the proof of Lemma B.1.1. Let ω be a $P\times 1$ nonrandom vector with $\|\omega\| = 1$.

$$\begin{aligned}
|\omega'c_{1,7}| &= \left| \frac{1}{NT}\sum_{i=1}^{N}\frac{\omega'X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\hat{F}'F\lambda_k\varepsilon_k'\varepsilon_i \right| \\
&= \left|\operatorname{tr}\left(\left(\frac{F'F}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{N}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\varepsilon_i\frac{\omega'X_i'F}{T}\right)\right| \\
&\lesssim \left\|\frac{1}{N}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\frac{\omega'X_i'}{\sqrt{T}}\right\| \\
&\leq \frac{1}{N}\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\right\|\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i\frac{\omega'X_i'}{\sqrt{T}}\right\| = O_p\left(\frac{1}{N}\right),
\end{aligned}$$

where the last equality holds by the fact that

$$E\left(\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\right\|^2\right) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i,j=1}^{N}E\left[\lambda_i'\lambda_j\right]E\left(\varepsilon_{it}\varepsilon_{jt}\right) \leq \max_{i,j}\left|E\left[\operatorname{tr}\left(\lambda_i'\lambda_j\right)\right]\right|\frac{1}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij} = O(1)$$

by Assumption B(i), B(iii) and C(ii), and similarly $E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i\omega'X_i'\right\|^2 = O_p(1)$. Note that the probability order of $c_{1,7}$ is the same as that in Bai (2009) and it is a potential bias term to be corrected. Therefore we denote $c_{1,7} = -C_{NT} = O_p\left(\frac{1}{N}\right)$. Lastly, for $c_{1,8}$ we have

$$c_{1,8} = \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}H'F'\varepsilon_k\varepsilon_k'\varepsilon_i + \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\left(\hat{F} - F^0H\right)'\varepsilon_k\varepsilon_k'\varepsilon_i$$
$$\equiv c_{1,81} + c_{1,82}.$$

Note that

$$c_{1,81} = \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}H'F'\varepsilon_k[\varepsilon_k'\varepsilon_i - E\left(\varepsilon_k'\varepsilon_i\right)] + \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}H'F'\varepsilon_k E\left(\varepsilon_k'\varepsilon_i\right)$$
$$\equiv c_{1,811} + c_{1,812}.$$

For $c_{1,811}$,

$$\begin{aligned}
|\omega' c_{1,811}| &= \left\|\frac{1}{N^2T^3}\sum_{i,k=1}^{N}F\tilde{G}H'F'\varepsilon_k[\varepsilon_k'\varepsilon_i - E\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i'\right\| \\
&\lesssim \left\|\frac{1}{N^2T^{5/2}}\sum_{i,k=1}^{N}F'\varepsilon_k[\varepsilon_k'\varepsilon_i - E\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i'\right\| \\
&\lesssim \frac{1}{T}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{NT}\sum_{i=1}^{N}[\varepsilon_k'\varepsilon_i - E\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i'\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{k=1}^{N}\|F'\varepsilon_k\|^2\right)^{\frac{1}{2}} \\
&= \frac{1}{T}O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon}\right)O_p\left(T^{d_\varepsilon}\right) = O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon - 1}\right),
\end{aligned}$$

where the last equality holds by the fact that

$$\begin{aligned}
&E\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{NT}\sum_{i=1}^{N}[\varepsilon_k'\varepsilon_i - E\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i'\right\|^2\right) \\
&= \frac{1}{N^3T}\sum_{i,j,k=1}^{N}E\left[\frac{\omega' X_i'X_j\omega}{T}\right]\sum_{t,s=1}^{T}E\left\{\left[\varepsilon_{it}\varepsilon_{kt} - E\left(\varepsilon_{it}\varepsilon_{kt}\right)\right]\left[\varepsilon_{js}\varepsilon_{ks} - E\left(\varepsilon_{js}\varepsilon_{ks}\right)\right]\right\} \\
&\leq \max_{i,j}E\left[\frac{\omega' X_i'X_j\omega}{T}\right]\frac{1}{N^3T}\sum_{i,j,k=1}^{N}\sum_{t,s=1}^{T}E\left[\left(\varepsilon_{kt}\varepsilon_{it} - E\left(\varepsilon_{kt}\varepsilon_{it}\right)\right)\left(\varepsilon_{ks}\varepsilon_{js} - E\left(\varepsilon_{ks}\varepsilon_{js}\right)\right)\right] \\
&\lesssim \frac{1}{N^3T}\sum_{i,j,k=1}^{N}\sum_{t,s=1}^{T}\left|\operatorname{cov}\left(\varepsilon_{it}\varepsilon_{kt}, \varepsilon_{js}\varepsilon_{ks}\right)\right| = O\left(N^{-1}T^{2d_\varepsilon}\right) \qquad \text{(B.2.3)}
\end{aligned}$$

by Assumption B(i), B(iii) and C(iv). Next,

$$\|c_{1,812}\| = \left\|\frac{1}{NT}\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\sum_{k=1}^{N}H'F'\varepsilon_k\frac{E\left(\varepsilon_k'\varepsilon_i\right)}{T}\right\| \lesssim \frac{1}{NT}\frac{1}{N}\sum_{i,k=1}^{N}\left\|\frac{X_i'F}{T}\right\|_F\|F'\varepsilon_k\|_F\,\overline{\sigma}_{ik},$$

where the expectation of the term is bounded above by

$$\frac{1}{NT}\max_{i,k} E\left(\left\|\frac{X_i'F}{T}\right\|\left\|F'\varepsilon_k\right\|_F\right)\frac{1}{N}\sum_{i,k=1}^{N}\overline{\sigma}_{ik}$$
$$\leq \frac{M}{N}\max_i\left(E\left\|\frac{X_i'F}{T}\right\|^2\right)^{\frac{1}{2}}\max_k\left(E\left\|\frac{F'\varepsilon_k}{T}\right\|^2\right)^{\frac{1}{2}} = O\left(N^{-1}T^{d_\varepsilon-\frac{1}{2}}\right).$$

So $c_{1,812} = O_p\left(N^{-1}T^{d_\varepsilon-1/2}\right)$ and $c_{1,81} = O_p(N^{-\frac{1}{2}}T^{2d_\varepsilon-1})$. For $c_{1,82}$, we make the following decomposition

$$c_{1,82} = \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\left(\hat{F}-F^0H\right)'\varepsilon_k\left[\varepsilon_k'\varepsilon_i - E\left(\varepsilon_k'\varepsilon_i\right)\right]$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\left(\hat{F}-F^0H\right)'\varepsilon_k E\left(\varepsilon_k'\varepsilon_i\right) \equiv c_{1,821}+c_{1,822}.$$

For the first term on the rhs, we have

$$\|c_{1,821}\| \lesssim \frac{1}{T}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\sum_{t=1}^{T}\left[\varepsilon_{kt}\varepsilon_{it}-E\left(\varepsilon_{kt}\varepsilon_{it}\right)\right]\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\left(\hat{F}-F^0H\right)'\varepsilon_k\right\|^2\right)^{\frac{1}{2}}$$
$$= T^{-1}O\left(N^{-1/2}T^{d_\varepsilon}\right)O_p\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right) = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-1}\left\|\hat{\delta}\right\|+N^{-1/2}T^{d_\varepsilon-1}\delta_{NT}^{-1}\right)$$

by (B.2.3) and the derivation of order of the term A_2 in the proof of Lemma B.1.2. In addition,

$$\|c_{1,822}\| \lesssim \frac{1}{NT}\frac{1}{N}\sum_{i,k=1}^{N}\left\|\frac{X_i'F}{T}\right\|\left\|\left(\hat{F}-F^0H\right)'\varepsilon_k\right\|_F\overline{\sigma}_{ik} = O_p\left(N^{-1}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$$

by Lemma B.1.1 and arguments as used to analyze $c_{1,812}$ above. Therefore we can conclude that

$$c_{1,82} = O_p\left(\left(N^{-\frac{1}{2}}T^{d_\varepsilon-1}+N^{-1}\right)\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right).$$

Lastly, we study d.

$$\|d\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}QF'\varepsilon_i\right\| \leq \frac{1}{T}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{X_i'F}{T}\right\|\left\|F'\varepsilon_i\right\|\left\|Q\right\|$$
$$\lesssim \frac{1}{T}\left\|Q\right\| = O_p\left(T^{-1}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$$

we we use the fact that $\|Q\| = O_p\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)$ derived above.

As in Bai (2009), the approximation error of the second part,

$$\frac{1}{NT}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\right)'\left(\mathbf{M}_F-\mathbf{M}_{\hat{F}}\right)\varepsilon_i \equiv \frac{1}{NT}\sum_{i=1}^{N}V_i'\left(\mathbf{M}_F-\mathbf{M}_{\hat{F}}\right)\varepsilon_i,$$

can be expressed by replacing X_i with V_i, and apply the same arguments and probability order as above. Then we concludes that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{\hat{F}}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{\hat{F}}\right]\varepsilon_i\\
&=\frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_F-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_F\right]\varepsilon_i-C_{NT}\\
&+O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)+\left\|\hat{\delta}\right\|^2+N^{-\frac{1}{2}}T^{2d_\varepsilon-1}\right).
\end{aligned}$$

This completes the proof of Lemma B.1.4. ■

Proof of Lemma B.1.5. (i) Note that

$$\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i}=\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}W_{\varepsilon,i}-\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\frac{\tilde{W}_F\tilde{W}_F^{*}}{T}W_{\varepsilon,i}\equiv A_1+A_2$$

under the restriction $\tilde{W}_F\in\mathcal{W}\equiv\left\{\tilde{W}_F:\tilde{W}_F^{*}\tilde{W}_F/T=\mathbb{I}_R\right\}$. We first study A_1. Recall that $X_{p,it}$ denote the p-th element of X_{it}. Let $W_{X_p,il}$ denote the p-th element of of $W_{X,il}$ and $W_{X_p,i}=(W_{X_p,i1},...,W_{X_p,iL})'$. The modulus of $\frac{1}{T}W_{X_p,i}^{*}W_{\varepsilon,i}$ satisfies

$$\begin{aligned}
\left|\frac{1}{T}W_{X_p,i}^{*}W_{\varepsilon,i}\right|&=\left|\frac{1}{T}\sum_{j=1}^{L}W_{X_p,ij}W_{\varepsilon,ij}^{*}\right|\leq\left|\frac{1}{T}\sum_{j=1}^{L}W_{X_p,ij}W_{X_p,ij}^{*}\right|^{\frac{1}{2}}\left|\frac{1}{T}\sum_{j=1}^{L}W_{\varepsilon,ij}W_{\varepsilon,ij}^{*}\right|^{\frac{1}{2}}\\
&\equiv\left|\hat{F}_{X_p,i}\left(\gamma_L\right)\right|^{\frac{1}{2}}\left|\hat{F}_{\varepsilon,i}\left(\gamma_L\right)\right|^{\frac{1}{2}}
\end{aligned}$$

by Cauchy-Schwarz inequality. Note that the two terms on the r.h.s. are averaged periodograms of $\{X_{p,it}\}_{t=1}^{T}$ and $\{\varepsilon_{it}\}_{t=1}^{T}$. Under Assumption A, A* and G, we can adopt Theorem 1 in Robinson (1994b) to obtain

$$\frac{\hat{F}_{X_p,i}\left(\gamma_L\right)}{F_{X_p,i}\left(\gamma_L\right)}\xrightarrow{p}1\text{ and }\frac{\hat{F}_{\varepsilon,i}\left(\gamma_L\right)}{F_{\varepsilon,i}\left(\gamma_L\right)}\xrightarrow{p}1\text{ as }T\to\infty,\tag{B.2.4}$$

where $F_{X_{p,i}}\left(\gamma_L\right)$ and $F_\varepsilon\left(\gamma_L\right)$ are the "pseudo spectral distribution" for $\{X_{p,it}\}_{t=1}^{T}$ and $\{\varepsilon_{it}\}_{t=1}^{T}$, respectively. Then we can conclude that $\hat{F}_{X_{p,i}}\left(\gamma_L\right)\sim\frac{\Upsilon_{i,XX,pp}}{1-2d_{X_p}}\gamma_L^{1-2d_{X_p}}$ and $\hat{F}_{\varepsilon,i}\left(\gamma_L\right)\sim\frac{\Upsilon_{i,\varepsilon}}{1-2d_\varepsilon}\gamma_L^{1-2d_\varepsilon}$, where $\Upsilon_{i,XX,pp}$ denote the (p,p)-th element of $\Upsilon_{i,XX}$. This result is compatible with our Assumption B*(i) and C*(i), and implies that for each p,

$$\left|\frac{1}{NT}\sum_{i=1}^{N}W_{X_p,i}^{*}W_{\varepsilon,i}\right| \lesssim \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\Upsilon_{i,XX,pp}}{1-2d_{X_p}}\gamma_L^{1-2d_{X_p}}\right|^{\frac{1}{2}}\left|\frac{\Upsilon_{i,\varepsilon\varepsilon}}{1-2d_{\varepsilon}}\gamma_L^{1-2d_{\varepsilon}}\right|^{\frac{1}{2}}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\Upsilon_{i,XX,pp}\Upsilon_{i,\varepsilon\varepsilon}}{\left(1-2d_{X_p}\right)\left(1-2d_{\varepsilon}\right)}\right)^{\frac{1}{2}}\gamma_L^{1-d_{X_p}-d_{\varepsilon}} = o_p(1)$$

by Assumption A*(i) and the fact that d_{X_p} and d_{ε} being strictly less than $\frac{1}{2}$. It follows that

$$\|A_1\|^2 = \sum_{p=1}^{P}\left|\frac{1}{NT}\sum_{i=1}^{N}W_{X_p,i}^{*}W_{\varepsilon,i}\right|^2 = O_p\left(\gamma_L^{2(1-d_{X,\max}-d_{\varepsilon})}\right) = o_p\left(1\right),$$

with $d_{X,\max} = \max_{1\leq p\leq P} d_{X_p}$. Next, for A_2 we have

$$A_2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}W_{X,i}^{*}\tilde{W}_F\right)\left(\frac{1}{T}\tilde{W}_F^{*}W_{\varepsilon,i}\right) \equiv \frac{1}{N}\sum_{i=1}^{N}A_{i,21}A_{i,22}.$$

Note that $A_{i,21}$ is a $P\times R$ matrix and $A_{i,22}$ an $R\times 1$ vector. Consider an arbitrary p-th element of $A_{i,21}A_{i,22}$, which is given by $\sum_{r=1}^{R}A_{i,21,pr}A_{i,22,r}$, where $A_{i,21,pr}$ and $A_{i,22,r}$ denotes the (p,r)-th element of $A_{i,21}$ and the rth element of $A_{i,22}$, respectively:

$$A_{i,21,pr} = \frac{1}{T}W_{X_p,i}^{*}\bar{W}_{F_r},\text{ and } A_{i,22,r} = \frac{1}{T}\bar{W}_{F_r}^{*}W_{\varepsilon,i},$$

where $W_{X_p,i}$ and $\bar{W}_{F_r}$ are both $L\times 1$ vectors that refer to the DFT of p-th element of the regressor X_{it} and the r-th element of F_t (which may not be the true vector in this lemma). By construction $\bar{W}_{F_r} = \gamma_L^{d_{F_r}-\frac{1}{2}}W_{F_r}$. Then using the same reasoning that analyzes A_1, we obtain $\left|\frac{1}{T}W_{X_p,i}^{*}\bar{W}_{F_r}\right| = O_p\left(\gamma_L^{\frac{1}{2}-d_{X_p}}\right)$ and $\left|\frac{1}{T}\tilde{W}_{F_r}^{*}W_{\varepsilon,i}\right| = O_p\left(\gamma_L^{\frac{1}{2}-d_{\varepsilon}}\right)$ uniformly in i. It follows that

$$\|A_2\|^2 = \sum_{p=1}^{P}\left|\sum_{r=1}^{R}\frac{1}{N}\sum_{i=1}^{N}A_{i,21,pr}A_{i,22,r}\right|^2 \lesssim \max_{p,r}\left|\frac{1}{N}\sum_{i=1}^{N}A_{i,21,pr}A_{i,22,r}\right|^2 = O_p\left(\gamma_L^{2(1-d_{X,\max}-d_{\varepsilon})}\right).$$

That is, $A_2 = O_p\left(\gamma_L^{1-d_{X,\max}-d_{\varepsilon}}\right)$. In sum, we have

$$\sup_{\tilde{W}_F\in\mathcal{W}}\left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i}\right\| = o_p\left(1\right).$$

(ii) and (iii): The proof is similar to that of (i) and thus omitted. ■

Proof of Lemma B.1.6. Let $\tilde{\delta} = \tilde{\beta}-\beta$. As in the proof of Lemma B.1.1, we consider the following eigenvalue problem

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)^{*}\right]\hat{W}_F=\hat{W}_F V_{NL}. \tag{B.2.5}$$

By expanding $W_{Y,i}$, (B.2.5) follows that

$$\begin{aligned}\hat{W}_F V_{NL}=&\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}\tilde{\delta}\tilde{\delta}'W_{X,i}^{*}\hat{W}_F-\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}\tilde{\delta}\lambda_i'W_{F^0}^{*}\hat{W}_F-\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}\tilde{\delta}W_{\varepsilon,i}^{*}\hat{W}_F\\&-\frac{1}{NT}\sum_{i=1}^{N}W_{F^0}\lambda_i\tilde{\delta}'W_{X,i}^{*}\hat{W}_F-\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}\tilde{\delta}'W_{X,i}^{*}\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}W_{F^0}\lambda_iW_{\varepsilon,i}^{*}\hat{W}_F\\&+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}\lambda_i'W_{F^0}^{*}\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}W_{\varepsilon,i}^{*}\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F^0}\tilde{\lambda}_i\tilde{\lambda}_i'\tilde{W}_{F^0}^{*}\hat{W}_F\\&\equiv\tilde{I}_1+\cdots+\tilde{I}_9.\end{aligned} \tag{B.2.6}$$

Since $\tilde{I}_9=\tilde{W}_{F^0}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)\left(\tilde{W}_{F^0}^{*}\hat{W}_F/T\right)$, we have

$$\hat{W}_F V_{NL}-\tilde{W}_{F^0}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)\left(\tilde{W}_{F^0}^{*}\hat{W}_F/T\right)=\tilde{I}_1+\cdots+\tilde{I}_8.$$

Recall that $\tilde{H}=\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\left(\frac{\tilde{W}_{F^0}^{*}\hat{W}_F}{T}\right)V_{NL}^{-1}$. Then

$$\begin{aligned}T^{-\frac{1}{2}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|&=T^{-\frac{1}{2}}\left\|\left[\hat{W}_F V_{NL}\left(\tilde{W}_{F^0}^{*}\hat{W}_F/T\right)^{-1}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)^{-1}-\tilde{W}_{F^0}\right]\tilde{H}\right\|\\&=T^{-\frac{1}{2}}\left\|\left(\tilde{I}_1+\cdots+\tilde{I}_8\right)\left(\tilde{W}_{F^0}^{*}\hat{W}_F/T\right)^{-1}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)^{-1}\tilde{H}\right\|\\&\lesssim T^{-\frac{1}{2}}\left(\left\|\tilde{I}_1\right\|+\cdots\left\|\tilde{I}_8\right\|\right)\|V_{NL}^{-1}\|\end{aligned}$$

given the invertibility of $\tilde{W}_{F^0}^{*}\hat{W}_F/T$ by Proposition 3.4.1 using the same reasoning as in the proof of Proposition 1 in Bai (2009). It is sufficient to study $\tilde{I}_1,\ldots,\tilde{I}_8$. For $\tilde{I}_1$, we have

$$\begin{aligned}T^{-\frac{1}{2}}\left\|\tilde{I}_1\right\|&=T^{-\frac{1}{2}}\left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\tilde{\delta}\tilde{\delta}'W_{X,i}^{*}\hat{W}_F\right\|\\&\leq\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\|W_{X,i}\|^2\left\|\tilde{\delta}\right\|^2T^{-\frac{1}{2}}\left\|\hat{W}_F\right\|\lesssim\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\|W_{X,i}\|^2\left\|\tilde{\delta}\right\|^2\\&=O_p\left(\gamma_L^{1-2d_{X,\max}}\right)\left\|\tilde{\delta}\right\|^2=o_p\left(\gamma_L^{1-2d_{X,\max}}\left\|\tilde{\delta}\right\|\right),\end{aligned}$$

where we use the fact that $T^{-\frac{1}{2}}\left\|\hat{W}_F\right\|=\sqrt{R}$ and $\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2=O_p\left(\gamma_L^{1-2d_{X,\max}}\right)$ by following arguments used in the proof of Lemma B.1.5 under Assumption B*(i). Similarly,

$$T^{-\frac{1}{2}}\left\|\tilde{I}_2\right\|=T^{-\frac{1}{2}}\left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}\tilde{\delta}\tilde{\lambda}_i'\tilde{W}_{F^0}^{*}\hat{W}_F\right\|$$

$$
\begin{aligned}
&\lesssim \frac{1}{NT^{\frac{1}{2}}} \sum_{i=1}^{N} \|W_{X,i}\| \left\|\tilde{\lambda}_i\right\| T^{-\frac{1}{2}} \left\|\tilde{W}_{F^0}\right\| \left\|\tilde{\delta}\right\| \\
&\lesssim \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|W_{X,i}\|^2 \right\}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\|\tilde{\lambda}_i\right\|^2 \right\}^{1/2} \left\|\tilde{\delta}\right\| \\
&= O_p\left(\gamma_L^{1/2-d_{X,\max}}\right) O_p\left(\gamma_L^{1/2-d_{F,\max}}\right) \left\|\tilde{\delta}\right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_{F,\max}} \left\|\tilde{\delta}\right\|\right)
\end{aligned}
$$

by Assumption B(iv), the fact that $\tilde{\lambda}_i = \tilde{\Gamma}_F^{-1}\lambda_i$ and that $\frac{1}{T}\left\|\tilde{W}_{F^0}\right\|^2 = O_p(1)$ by following arguments used in the proof of Lemma B.1.5. Analogously, we have

$$
T^{-\frac{1}{2}} \left\|\tilde{I}_3\right\| = T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{X,i} \tilde{\delta} W_{\varepsilon,i}^{*} \hat{W}_F \right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon} \left\|\tilde{\delta}\right\|\right),
$$

$$
T^{-\frac{1}{2}} \left\|\tilde{I}_4\right\| = T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \tilde{W}_{F^0} \tilde{\lambda}_i \tilde{\delta}' W_{X,i}^{*} \hat{W}_F \right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_F} \left\|\tilde{\delta}\right\|\right),
$$

and

$$
T^{-\frac{1}{2}} \left\|\tilde{I}_5\right\| = T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{\varepsilon,i} \tilde{\delta}' W_{X,i}^{*} \hat{W}_F \right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon} \left\|\tilde{\delta}\right\|\right).
$$

For $\tilde{I}_6$ we have

$$
\begin{aligned}
T^{-\frac{1}{2}} \left\|\tilde{I}_6\right\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \tilde{W}_{F^0} \tilde{\lambda}_i W_{\varepsilon,i}^{*} \hat{W}_F \right\| \lesssim \frac{1}{NT^{\frac{1}{2}}} \left\| \sum_{i=1}^{N} \tilde{\lambda}_i W_{\varepsilon,i}^{*} \right\| \\
&= O_p\left(N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon}\right),
\end{aligned}
$$

by Assumption C*(i) and C*(ii) using the similar reasoning as in the proof of Lemma 1(ii) in Bai and Ng (2002). Note that $\tilde{I}_7$ is a conjugate transpose of $\tilde{I}_6$, so $T^{-\frac{1}{2}}\left\|\tilde{I}_7\right\| = T^{-\frac{1}{2}}\left\|\tilde{I}_6\right\|$ and share the same order. For $\tilde{I}_8$, we follow the reasoning as used in the proof of Lemma B.1.1 and consider the transpose of the l-th row of $\tilde{I}_8$ as

$$
\begin{aligned}
\tilde{I}_{8,l} &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{k=1}^{L} W_{\varepsilon,il} W_{\varepsilon,ik}^{*} \hat{W}_{F,k} \\
&= \frac{1}{NT} \sum_{i=1}^{N} \sum_{k=1}^{L} E\left(W_{\varepsilon,il} W_{\varepsilon,ik}^{*}\right) \hat{W}_{F,k} + \frac{1}{T} \sum_{k=1}^{L} \left(\frac{1}{N} \sum_{i=1}^{N} W_{\varepsilon,il} W_{\varepsilon,ik}^{*} - \gamma_N^W(k,l) \right) \hat{W}_{F,k} \equiv \tilde{I}_{8,l1} + \tilde{I}_{8,l2},
\end{aligned}
$$

where $\gamma_N^W(k,l) = \frac{1}{N}\sum_{i=1}^{N} E\left(W_{\varepsilon,il} W_{\varepsilon,ik}^{*}\right)$ and $\hat{W}_{F,k}$ denotes the k-th column of $\hat{W}_F$. Note that

$$
\begin{aligned}
\frac{1}{T} \sum_{l=1}^{L} \left\|\tilde{I}_{8,l1}\right\|^2 &= \frac{1}{N^2T^2} \sum_{l=1}^{L} \left\| \sum_{i=1}^{N} \sum_{k=1}^{L} E\left(W_{\varepsilon,il} W_{\varepsilon,ik}^{*}\right) \hat{W}_{F,k} \right\|^2 \\
&= \frac{1}{N^2T^3} \sum_{l=1}^{L} \sum_{i,j=1}^{N} \sum_{k,m=1}^{L} E\left(W_{\varepsilon,il} W_{\varepsilon,ik}^{*}\right) E\left(W_{\varepsilon,jl}^{*} W_{\varepsilon,jm}\right) \hat{W}_{F,k}^{*} \hat{W}_{F,m}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{N^2T^3}\sum_{l=1}^{L}\sum_{i,j=1}^{N}\sum_{k,m=1}^{L}\sqrt{E\left|W_{\varepsilon,il}W_{\varepsilon,ik}^{*}\right|^2 E\left|W_{\varepsilon,jl}^{*}W_{\varepsilon,jm}\right|^2}\left\|\hat{W}_{F,k}\right\|\left\|\hat{W}_{F,m}\right\| \\
&\leq \frac{1}{N^2T^3}\sum_{l=1}^{L}\sum_{i,j=1}^{N}\sum_{k,m=1}^{L}\gamma_l^{-2d_\varepsilon}\gamma_k^{-d_\varepsilon}\gamma_m^{-d_\varepsilon}\overline{\sigma}_{ij}^{W}\left\|\hat{W}_{F,k}\right\|\left\|\hat{W}_{F,m}\right\| \\
&\lesssim \frac{1}{NT^3}\left(\sum_{l=1}^{L}\gamma_l^{-2d_\varepsilon}\right)^2\left(\sum_{l=k}^{L}\left\|\hat{W}_{F,k}\right\|^2\right) = O_p\left(\frac{1}{N}\gamma_L^{2-4d_\varepsilon}\right)
\end{aligned}
$$

by Assumption C*(ii). In addition, $\frac{1}{T}\sum_{l=1}^{L}\left\|\tilde{I}_{8,l2}\right\|^2 = O_p\left(N^{-1}\gamma_L^{2-4d_\varepsilon}\right)$ following the same reasoning as above and using Assumption C*(iv). In sum, we have

$$
\begin{aligned}
T^{-\frac{1}{2}}\left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\| &= O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right) \\
&\equiv O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right),
\end{aligned}
$$

which completes the proof of Lemma B.1.6. ■

Proof of Lemma B.1.7. The proof of this lemma parallels that of Lemma B.1.2. Note that

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\hat{W}_F}{T} &= \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\tilde{W}_{F^0}\tilde{H}}{T} + \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)}{T} \\
&\equiv A_1 + A_2. \qquad \text{(B.2.7)}
\end{aligned}
$$

For A_1, we have

$$
\begin{aligned}
\|A_1\| &\leq \left\|\tilde{H}\right\|\frac{1}{NT}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\tilde{W}_{F^0}\right\| \\
&= O_p\left(\gamma_L^{1-2d_{F,\max}}\right)O_p\left(N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right) = O_p\left(N^{-\frac{1}{2}}N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right)
\end{aligned}
$$

by B*(iii) and D*(iii), and the fact that $\tilde{H} = O_p\left(\gamma_L^{1-2d_{F,\max}}\right)$ as in the proof of Lemma B.1.6. Next we denote

$$
\tilde{G} = \left(\frac{\tilde{W}_{F^0}^{*}\hat{W}_F}{T}\right)^{-1}\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1} = O_p\left(\gamma_L^{2d_{F,\min}-1}\right).
$$

Following the proof of Lemma B.1.6, we have for A_2 that

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)}{T} &= \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\left(\hat{W}_F\tilde{H}^{-1} - \tilde{W}_{F^0}\right)}{T}\tilde{H} \\
&= \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\left(\tilde{I}_1 + \cdots + \tilde{I}_8\right)\tilde{G}}{T}\tilde{H}
\end{aligned}
$$

$$\lesssim \frac{1}{NT}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\left(\tilde{I}_1+\cdots+\tilde{I}_8\right)\equiv a_1+\cdots+a_8,$$

using the fact that $\left\|\tilde{G}\tilde{H}\right\| = \|V_{NL}^{-1}\| = O_p(1)$. For a_1, we have

$$\begin{aligned}
\|a_1\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}\tilde{\delta}\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}^{*}\|^2\left\|\tilde{\delta}\right\|^2 \\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1-2d_{X,\max}}\right)\left\|\tilde{\delta}\right\|^2 = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{X,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|^2\right)
\end{aligned}$$

by Assumption B*(i) and C*(i). Similarly,

$$\begin{aligned}
\|a_2\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}\tilde{\delta}\lambda_k'W_{F^0}^{*}\hat{W}_F\right\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\|W_{X,k}\|\|\lambda_k\|\left\|\tilde{\delta}\right\|T^{-\frac{1}{2}}\|W_{F^0}\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2\right\}^{1/2}T^{-\frac{1}{2}}\|W_{F^0}\|\left\|\tilde{\delta}\right\| \\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1/2-d_{X,\max}}\right)O_p\left(\gamma_L^{1/2-d_{F,\max}}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{F,\max}-d_{X,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|\right),
\end{aligned}$$

$$\begin{aligned}
\|a_3\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}\tilde{\delta}W_{\varepsilon,k}^{*}\hat{W}_F\right\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|\|W_{\varepsilon,k}\|\left\|\tilde{\delta}\right\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{\varepsilon,k}\|^2\right\}^{1/2}\left\|\tilde{\delta}\right\| \\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1/2-d_{X,\max}}\right)O_p\left(\gamma_L^{1/2-d_\varepsilon}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-2d_\varepsilon}\left\|\tilde{\delta}\right\|\right),
\end{aligned}$$

$$\begin{aligned}
\|a_4\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{F^0}\lambda_k\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\|W_{X,k}\|\|\lambda_k\|T^{-\frac{1}{2}}\|W_{F^0}\|\left\|\tilde{\delta}\right\| \\
&\leq \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2\right\}^{1/2}T^{-\frac{1}{2}}\|W_{F^0}\|\left\|\tilde{\delta}\right\| \\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1/2-d_{X,\max}}\right)O_p\left(\gamma_L^{1/2-d_{F,\max}}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-d_{F,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|\right),
\end{aligned}$$

and

$$\|a_5\| = \left\| \frac{1}{N}\sum_{i=1}^{N}\lambda_i \frac{1}{T} W_{\varepsilon,i}^* \frac{1}{NT}\sum_{k=1}^{N} W_{\varepsilon,k}\tilde{\delta}' W_{X,k}^* \hat{W}_F \right\|$$
$$\lesssim \frac{1}{NT^{\frac{1}{2}}} \left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^*\right\| \left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{\varepsilon,k}\|^2\right\}^{1/2} \left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2\right\}^{1/2} \left\|\tilde{\delta}\right\|$$
$$= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right) O_p\left(\gamma_L^{1/2-d_\varepsilon}\right) O_p\left(\gamma_L^{1/2-d_{X,\max}}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-2d_\varepsilon}\left\|\tilde{\delta}\right\|\right).$$

For a_6 we make the following decomposition

$$a_6 = \frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^* \sum_{k=1}^{N} W_{F^0}\lambda_k W_{\varepsilon,k}^* \left[\tilde{W}_{F^0}\tilde{H} + \left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\right] \equiv a_{6,1} + a_{6,2}.$$

Note that

$$\|a_{6,1}\| = \left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^* W_{F^0}\sum_{k=1}^{N}\lambda_k W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H}\right\| \le \left\|\tilde{\Gamma}_F\tilde{H}\right\| \left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^* W_{F^0}\right\|^2$$
$$\lesssim \gamma_L^{\frac{1}{2}-d_{F,\max}} \left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^* W_{F^0}\right\|^2$$
$$= O_p\left(\gamma_L^{\frac{1}{2}-d_{F,\max}}\right) O_p\left(N^{-1}L^{-1}\gamma_L^{2-2d_\varepsilon}\right) O_p\left(\gamma_L^{-2d_{F,\max}}\right) = O_p\left(N^{-1}L^{-1}\gamma_L^{\frac{5}{2}-3d_{F,\max}-2d_\varepsilon}\right),$$

by Assumption D*(iii). Next

$$\|a_{6,2}\| = \left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^* W_{F^0}\sum_{k=1}^{N}\lambda_k W_{\varepsilon,k}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\right\|$$
$$\le \frac{1}{N^2T}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^*\right\|^2 T^{-\frac{1}{2}}\|W_{F^0}\| T^{-\frac{1}{2}}\left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\|$$
$$= O_p\left(N^{-1}\gamma_L^{1-2d_\varepsilon}\right) O_p\left(\gamma_L^{\frac{1}{2}-d_{F,\max}}\right) O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)$$
$$= O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right),$$

by Lemma B.1.1, where $\delta_{W1,NT} = \gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$. And the order of $\left\|\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^*\right\|^2$ above is obtained because

$$E\left\|\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^*\right\|^2 \frac{1}{T} = \frac{1}{N^2T}\sum_{i,k=1}^{N} E\left(\lambda_k'\lambda_i\right) E\left(W_{\varepsilon,i}^* W_{\varepsilon,k}\right)$$
$$\le \max_{i,k}\left|E\left(\lambda_k'\lambda_i\right)\right| \frac{1}{N^2T}\sum_{i,k=1}^{N}\sum_{l=1}^{L}\left|E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right| = O\left(N^{-1}\gamma_L^{1-2d_\varepsilon}\right)$$

by Assumption B(iv), C*(i) and C*(ii). Then

$$a_6 = O_p\left(N^{-1}L^{-1}\gamma_L^{\frac{5}{2}-3d_{F,\max}-2d_\varepsilon} + N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right).$$

For a_7 we have, by the same reasoning as $a_{6,2}$ that

$$\begin{aligned}
\|a_7\| &= \left\|\frac{1}{N}\sum_{i=1}^N \lambda_i \frac{1}{T} W_{\varepsilon,i}^* \left(\frac{1}{NT}\sum_{k=1}^N W_{\varepsilon,k}\lambda_k' W_{F^0}^* \hat{W}_F\right)\right\| \\
&\lesssim \frac{1}{N^2 T}\left\|\sum_{i=1}^N \lambda_i W_{\varepsilon,i}^*\right\|^2 T^{-\frac{1}{2}}\|W_{F^0}\| \\
&= O_p\left(N^{-1}\gamma_L^{1-2d_\varepsilon}\right) O_p\left(\gamma_L^{\frac{1}{2}-d_{F,\max}}\right) = O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\right).
\end{aligned}$$

For a_8, we have

$$a_8 = \frac{1}{N^2T^2}\sum_{i=1}^N \lambda_i W_{\varepsilon,i}^* \sum_{k=1}^N W_{\varepsilon,k} W_{\varepsilon,k}^* \left[\tilde{W}_{F^0}\tilde{H} + \left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\right] \equiv a_{8,1} + a_{8,2}.$$

Note that

$$\begin{aligned}
a_{8,1} &= \frac{1}{N^2T^2}\sum_{i=1}^N\sum_{k=1}^N \lambda_i \sum_{l=1}^L \left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right] W_{\varepsilon,k}^* \tilde{W}_{F^0}\tilde{H} \\
&+ \frac{1}{N^2T^2}\sum_{i=1}^N\sum_{k=1}^N \lambda_i \sum_{l=1}^L E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right) W_{\varepsilon,k}^* \tilde{W}_{F^0}\tilde{H} \equiv a_{8,11} + a_{8,12}.
\end{aligned}$$

For $a_{8,11}$, we have

$$\begin{aligned}
\|a_{8,11}\| &\le \frac{1}{N^2T^2}\sum_{k=1}^N \left\|\sum_{i=1}^N \lambda_i \sum_{l=1}^L \left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right]\right\| \left\|W_{\varepsilon,k}^* \tilde{W}_{F^0}\tilde{H}\right\| \\
&\le \frac{1}{\sqrt{N}T}\left(\frac{1}{N^2}\sum_{k=1}^N \left\|\sum_{i=1}^N \lambda_i \sum_{l=1}^L \left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right]\right\|^2\right)^{\frac{1}{2}} \left(\frac{1}{NT^2}\sum_{k=1}^N \left\|W_{\varepsilon,k}^* \tilde{W}_{F^0}\tilde{H}\right\|^2\right)^{\frac{1}{2}} \\
&= O\left(N^{-\frac{1}{2}}T^{-1}\right) O_p\left(L\gamma_L^{-2d_\varepsilon}\right) O_p\left(L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right) \\
&= O_p\left(N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{5}{2}-2d_{F,\max}-3d_\varepsilon}\right)
\end{aligned}$$

by Assumption B(iv), B*(iii), C*(i), C*(v) and E*(iii) using the same reasoning as studying a_{81a} in the proof of Lemma B.1.2. To be specific,

$$\begin{aligned}
&E\left[\frac{1}{N^2}\sum_{k=1}^N \left\|\sum_{i=1}^N \lambda_i \sum_{l=1}^L \left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right]\right\|^2\right] \\
&= \frac{1}{N^2}\sum_{h,i,k=1}^N E\left(\lambda_i^* \lambda_h\right) \sum_{l,m=1}^L E\left\{\left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right]\left[W_{\varepsilon,hm}^* W_{\varepsilon,km} - E\left(W_{\varepsilon,hm}^* W_{\varepsilon,km}\right)\right]\right\}
\end{aligned}$$

$$= \frac{1}{N^2} \sum_{h,i,k=1}^{N} E\left(\lambda_i^* \lambda_h\right) \sum_{l,m=1}^{L} \operatorname{cov}\left[W_{\varepsilon,il}^* W_{\varepsilon,kl}, W_{\varepsilon,hm}^* W_{\varepsilon,km}\right]$$
$$\leq \frac{\max_{i,h} E\left(\lambda_i^* \lambda_h\right)}{N^2} \sum_{h,i,k=1}^{N} \sum_{l,m=1}^{L} \left|\operatorname{cov}\left[W_{\varepsilon,il}^* W_{\varepsilon,kl}, W_{\varepsilon,hm}^* W_{\varepsilon,km}\right]\right| = O\left(L^2 \gamma_L^{-4d_\varepsilon}\right).$$

Next, $a_{8,12}$ follows that

$$\|a_{8,12}\| \leq \frac{1}{N^2T^2} \sum_{i=1}^{N}\sum_{k=1}^{N} \|\lambda_i\| \left(\sum_{l=1}^{L} \left|E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right|\right) \left\|W_{\varepsilon,k}^* \tilde{W}_{F^0} \tilde{H}\right\|$$
$$\lesssim \gamma_L^{1-2d_{F,\max}} \frac{1}{N^2 T} \sum_{i=1}^{N} \|\lambda_i\| \sum_{k=1}^{N} \left(\sum_{l=1}^{L} \left|E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right|\right) \frac{1}{\sqrt{T}} \left\|W_{\varepsilon,k} \tilde{W}_{F^0}\right\|$$
$$\lesssim \frac{\gamma_L^{1-2d_{F,\max}}}{\sqrt{N}T} \left(\frac{1}{N}\sum_{i=1}^{N} \|\lambda_i\|^2\right)^{\frac{1}{2}} \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} \left(\sum_{l=1}^{L} \left|E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right|\right)^2\right)^{\frac{1}{2}} \left(\frac{1}{N}\sum_{k=1}^{N} \frac{1}{T} \left\|W_{\varepsilon,k} \tilde{W}_{F^0}\right\|^2\right)^{\frac{1}{2}}$$
$$= O_p\left(\gamma_L^{1-2d_{F,\max}} N^{-\frac{1}{2}} T^{-1}\right) O_p(1)\, O\left(L\gamma_L^{-2d_\varepsilon}\right) O_p\left(L^{-\frac{1}{2}} \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$$
$$= O_p\left(N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{5}{2}-2d_{F,\max}-3d_\varepsilon}\right)$$

by Cauchy-Schwarz inequality and Assumptions B(iv), C*(i), C*(ii) and D*(iii). To be specific,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} \left(\sum_{l=1}^{L} \left|E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right|\right)^2 \leq \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} \left(\sum_{l=1}^{L} \sqrt{E\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)^2}\right)^2$$
$$\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} \left(\bar{\sigma}_{ij}^W\right)^2 \left(\sum_{l=1}^{L} \gamma_l^{-2d_\varepsilon}\right)^2 = O\left(L^2 \gamma_L^{-4d_\varepsilon}\right)$$

Therefore, $a_{8,1} = O_p\left(N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{5}{2}-2d_{F,\max}-3d_\varepsilon}\right)$. As for the order of $a_{8,2}$, the similar reasoning holds except we replace $T^{-\frac{1}{2}} \left\|\tilde{W}_{F^0}\tilde{H}\right\|$ by $T^{-\frac{1}{2}} \left\|\hat{W}_F - \tilde{W}_{F^0} H\right\|$, therefore

$$\|a_{8,2}\| = O_p\left(N^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-3d_\varepsilon} \left(\delta_{W1,NT} \left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right),$$

where $\delta_{W1,NT}$ is defined in Lemma B.1.6. Then

$$a_8 = O_p\left(N^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-3d_\varepsilon} \left(L^{-\frac{1}{2}} \gamma_L^{1-2d_{F,\max}} + \delta_{W1,NT} \left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right).$$

In sum, we have

$$\left\|\frac{1}{N}\sum_{i=1}^{N} \lambda_i \frac{W_{\varepsilon,i}^* \hat{W}_F}{T}\right\| = O_p\left(N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right) + O_p\left(N^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{X,\max}-d_\varepsilon} \left\|\tilde{\delta}\right\|^2\right)$$
$$+ O_p\left(N^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-d_{X,\max}-d_{F,\max}-d_\varepsilon} \left\|\tilde{\delta}\right\|\right) + O_p\left(N^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-d_{X,\max}-2d_\varepsilon} \left\|\tilde{\delta}\right\|\right)$$

$$
\begin{aligned}
&+ O_p\left(N^{-1}L^{-1}\gamma_L^{\frac{5}{2}-3d_{F,\max}-2d_\varepsilon}\right) + O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\right) \\
&+ O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right) \\
&+ O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-3d_\varepsilon}\left(L^{-\frac{1}{2}}\gamma_L^{1-2d_{F,\max}} + \delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right) \\
&\equiv O_p\left(\delta_{W,NL}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right),
\end{aligned}
$$

where $\delta_{W,NL} = N^{-\frac{1}{2}}\gamma_L^{1-d_{X,\max}-d_\varepsilon}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$. This completes the proof of Lemma B.1.7. ■

Proof of Lemma B.1.8. For $\tilde{J}_8$ we have the following decomposition:

$$
\begin{aligned}
\tilde{J}_8 &= -\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\right) \\
&= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\right) + \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\hat{W}_F^*W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\right) \\
&\equiv \tilde{J}_{8,a} + \tilde{J}_{8,b}.
\end{aligned}
$$

Firstly $\tilde{J}_{8,a}$ can be decomposed similarly into

$$
\begin{aligned}
&\tilde{J}_{8,a} \\
&= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H}\check{G}\lambda_i\right) - \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right) \\
&= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\sum_{j=1}^{L}\overline{W}_{X,ij}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}_{F^0,j}'\tilde{H}\check{G}\lambda_i\right) - \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\sum_{j\neq l}^{L}\overline{W}_{X,ij}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kl}\tilde{W}_{F^0,l}'\tilde{H}\check{G}\lambda_i\right) \\
&- \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right) \\
&\equiv \tilde{J}_{8,a1} + \tilde{J}_{8,a2} + \tilde{J}_{8,a3},
\end{aligned}
$$

and then $\tilde{J}_{8,b}$ could have the similar decomposition given by

$$
\begin{aligned}
\tilde{J}_{8,b} &= \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\tilde{H}^*\sum_{j=1}^{L}\overline{\tilde{W}}_{F^0,j}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}_{F^0,j}'\tilde{H}\check{G}\lambda_i\right) \\
&+ \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\tilde{H}^*\sum_{j\neq l}\overline{\tilde{W}}_{F^0,j}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kl}\tilde{W}_{F^0,l}'\tilde{H}\check{G}\lambda_i\right) \\
&+ \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\hat{W}_F^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right) \\
&+ \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)
\end{aligned}
$$

$$\equiv \tilde{J}_{8,b1} + \tilde{J}_{8,b2} + \tilde{J}_{8,b3} + \tilde{J}_{8,b4}.$$

Next, define $\tilde{J}_{8,1} = \tilde{J}_{8,a1} + \tilde{J}_{8,b1}$, and define $\tilde{J}_{8,2}$ and $\tilde{J}_{8,3}$ in the same manner. It is easy to see that

$$\begin{aligned}
\tilde{J}_{8,1} &= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\sum_{j=1}^{L}\overline{W}_{X,ij}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}'_{F^0,j}\tilde{H}\check{G}\lambda_i\right)\\
&+\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\frac{W^*_{X,i}\hat{W}_F}{T}\tilde{H}^*\sum_{j=1}^{L}\overline{\tilde{W}}_{F^0,j}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}'_{F^0,j}\tilde{H}\check{G}\lambda_i\right)\\
&= -\frac{1}{NT}\sum_{i=1}^{N}\text{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}Diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\check{G}\lambda_i\right) = A_{NT}.
\end{aligned}$$

Suppose we define $J_{it} = \left(X'_{it}, F^{0\prime}_t\right)'$, and $W_{J,ij}$ its DFT at frequency γ_j. Then $\tilde{J}_{8,a1}$ and $\tilde{J}_{8,b1}$ correspond to the submatrices of $\sum_{j=1}^{L}\overline{W}_{J,ij}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}'_{J,ij}$ with different weighted sum over i and k. And the same notation works for $\tilde{J}_{8,2}$. Following (22)-(35) in the proof of Theorem 2 in Christensen and Nielsen (2006) and using Cauchy-Schwarz inequality, which is reflected by our Assumption D(iii), it can be concluded that $\tilde{J}_{8,1} = O_p\left(\frac{1}{L}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$ and $\tilde{J}_{8,2} = O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$.

Then we analyze the order of $\tilde{J}_{8,3}$, which follows that

$$\begin{aligned}
\tilde{J}_{8,3} &= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{\varepsilon,k}W^*_{\varepsilon,k}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)\\
&= -\frac{1}{NT}\sum_{i=1}^{N}\text{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}E\left(W_{\varepsilon,k}W^*_{\varepsilon,k}\right)\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)\\
&-\frac{1}{NT}\sum_{i=1}^{N}\text{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}\left(W_{\varepsilon,k}W^*_{\varepsilon,k} - E\left(W_{\varepsilon,k}W^*_{\varepsilon,k}\right)\right)\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)\\
&\equiv \tilde{J}_{8,31} + \tilde{J}_{8,32},
\end{aligned}$$

where $\tilde{J}_{8,31}$ is given in norm by

$$\begin{aligned}
\left\|\tilde{J}_{8,31}\right\| &\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\frac{1}{NT}\sum_{i=1}^{N}\left\|\frac{W_{X,i}}{\sqrt{T}}\right\|\|\lambda_i\|\frac{1}{N}\sum_{k=1}^{N}\|\Omega_k\|_{sp}\left\|\check{\Gamma}_\varepsilon^{-1}\right\|_{sp}^2\frac{1}{\sqrt{T}}\left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\|\\
&= O_p\left(T^{2d_\varepsilon-1}\gamma_L^{d_{F,\min}-d_{X,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)
\end{aligned}$$

by Assumption C*(iii). And for $\tilde{J}_{8,32}$ we have

$$\begin{aligned}
\left\|\tilde{J}_{8,32}\right\| &\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|W_{X,i}\|\|\lambda_i\|\frac{1}{NT}\left\|\sum_{k=1}^{N}\left(W_{\varepsilon,k}W^*_{\varepsilon,k} - E\left(W_{\varepsilon,k}W^*_{\varepsilon,k}\right)\right)\right\|\frac{1}{\sqrt{T}}\left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\|\\
&= O_p\left(\gamma_L^{d_{F,\min}-\frac{1}{2}}\right)O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right)O_p\left(N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)
\end{aligned}$$

$$= O_p\left(N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon+(d_{F,\min}-d_{X,\max})}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right),$$

because

$$E\left[\frac{1}{NT}\left\|\sum_{k=1}^{N}\left(W_{\varepsilon,k}W_{\varepsilon,k}^*-E\left(W_{\varepsilon,k}W_{\varepsilon,k}^*\right)\right)\right\|\right]$$
$$\leq \frac{1}{NT}\left(E\sum_{i,k=1}^{N}\operatorname{tr}\left[\left(W_{\varepsilon,i}W_{\varepsilon,i}^*-E\left(W_{\varepsilon,i}W_{\varepsilon,i}^*\right)\right)\left(W_{\varepsilon,k}W_{\varepsilon,k}^*-E\left(W_{\varepsilon,k}W_{\varepsilon,k}^*\right)\right)^*\right]\right)^{\frac{1}{2}}$$
$$= \frac{1}{NT}\left(\sum_{i,k=1}^{N}\sum_{l,m=1}^{L}\operatorname{cov}\left(W_{\varepsilon,il}W_{\varepsilon,im}^*,W_{\varepsilon,kl}W_{\varepsilon,km}^*\right)\right)^{\frac{1}{2}} = O\left(N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right)$$

by Assumption C*(v). And the order of $\tilde{J}_{8,b4}$ can be omitted as it is dominated by $\tilde{J}_{8,b3}$. Thus we complete the proof of Lemma B.1.8. ■

Proof of Lemma B.1.9. We first consider how $\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right)$ converges to

$$\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^*\mathbf{M}_{W_{F^0}}W_{\varepsilon,i}\right).$$

Noting that $\mathbf{M}_{\hat{W}_F}-\mathbf{M}_{W_{F^0}}=\mathbf{P}_{W_{F^0}}-\mathbf{P}_{\hat{W}_F}$ and $\mathbf{P}_{\hat{W}_F}=\hat{W}_F\hat{W}_F^*/T$, we have

$$\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left[W_{X,i}^*\left(\mathbf{P}_{\hat{W}_F}-\mathbf{P}_{W_{F^0}}\right)W_{\varepsilon,i}\right]$$
$$=\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^*\frac{\hat{W}_F\hat{W}_F^*}{T}W_{\varepsilon,i}-W_{X,i}^*\mathbf{P}_{\tilde{W}_{F^0}}W_{\varepsilon,i}\right)$$
$$=\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^*\frac{\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^*\tilde{W}_{F^0}^*}{T}W_{\varepsilon,i}\right)$$
$$+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^*\frac{\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*}{T}W_{\varepsilon,i}\right)$$
$$+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^*\frac{\tilde{W}_{F^0}\tilde{H}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*}{T}W_{\varepsilon,i}\right)$$
$$+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\left(\tilde{H}\tilde{H}^*-\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\right)\tilde{W}_{F^0}^*W_{\varepsilon,i}\right)$$
$$\equiv a+b+c+d.$$

We study a, b, c and d in turn. First, for a we have

$$\begin{aligned}
\|a\| &= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \frac{\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^*\tilde{W}_{F^0}^*}{T} W_{\varepsilon,i}\right)\right\| \\
&\leq \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|}{NT^{\frac{3}{2}}} \sum_{i=1}^{N} \|W_{X,i}\| \left\|\tilde{W}_{F^0}^* W_{\varepsilon,i}\right\| \left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\| \frac{1}{T^{\frac{1}{2}}} \left\|\tilde{H}\right\| \\
&\lesssim \gamma_L^{1-2d_{F,\max}} \sqrt{NL}\gamma_L^{d_\varepsilon}\gamma_L^{-1} \|\Gamma_Z\| \left\{ \frac{1}{NT^{\frac{3}{2}}} \sum_{i=1}^{N} \|W_{X,i}\| \left\|\tilde{W}_{F^0}^* W_{\varepsilon,i}\right\| \right\} \frac{1}{T^{\frac{1}{2}}} \left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\| \\
&= \gamma_L^{1-2d_{F,\max}} \sqrt{NL}\gamma_L^{d_\varepsilon}\gamma_L^{d_{Z,\min}-1} O_p\left(L^{-\frac{1}{2}}\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right) O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right) \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1} O_p\left(\gamma_L^{2d_{Z,\min}-2d_{F,\max}+1-d_{X,\max}-d_\varepsilon}\delta_{W1,NT}\left\|\tilde{\delta}\right\|\right) \\
&+ O_p\left(\gamma_L^{2-2d_{F,\max}-2d_\varepsilon+d_{Z,\min}-d_{X,\max}}\right) \\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\| \left\|\tilde{\delta}\right\|\right) + o_p(1),
\end{aligned}$$

by Lemma B.1.6 and the fact that

$$\left\|\tilde{H}\right\| = \left\|\tilde{\Gamma}_F^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)\tilde{\Gamma}_F^{-1}\left(\frac{\tilde{W}_{F^0}^*\hat{W}_F}{T}\right)V_{NL}^{-1}\right\| = O_p\left(\gamma_L^{1-2d_{F,\max}}\right),$$

and an order of L small enough relative to T that represents an undersmoothed estimator.

Next, for b we have

$$\begin{aligned}
\|b\| &= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \frac{\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)^*}{T} W_{\varepsilon,i}\right)\right\| \\
&\leq \sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\| \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|W_{X,i}\| \|W_{\varepsilon,i}\| \right\} \frac{1}{T} \left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\|_F^2 \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon-1}\gamma_L^{d_{Z,\min}} O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right) O_p\left(\delta_{W1,NT}^2\left\|\tilde{\delta}\right\|^2 + N^{-1}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon}\right) \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1} O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}-d_\varepsilon}\delta_{W1,NT}^2\left\|\tilde{\delta}\right\|^2\right) + O_p\left(\sqrt{\frac{L}{N}}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon+d_{Z,\min}-d_{X,\max}}\right) \\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\| \left\|\tilde{\delta}\right\|\right) + o_p(1),
\end{aligned}$$

as it relies on some milder conditions relative to $\|a\|$.

Next, for c we have

$$\begin{aligned}
c &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^* \frac{\tilde{W}_{F^0}\tilde{H}\tilde{H}^*\left(\hat{W}_F\tilde{H}^{-1} - \tilde{W}_{F^0}\right)^*}{T} W_{\varepsilon,i}\right) \\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(\frac{W_{X,i}^*\tilde{W}_{F^0}}{T} \left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1} \left(\hat{W}_F\tilde{H}^{-1} - \tilde{W}_{F^0}\right)^* W_{\varepsilon,i}\right)
\end{aligned}$$

$$+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\tilde{Q}\left(\hat{W}_F\tilde{H}^{-1}-\tilde{W}_{F^0}\right)^{*}W_{\varepsilon,i}\right)\equiv c_1+c_2.$$

where $\tilde{Q}=\tilde{H}\tilde{H}^{*}-\left(\frac{\tilde{W}_{F^0}^{*}\tilde{W}_{F^0}}{T}\right)^{-1}$. For c_2, we have

$$\begin{aligned}
\|c_2\| &= \left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\tilde{Q}\tilde{H}^{*-1}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}W_{\varepsilon,i}\right\|\\
&\leq \sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\left\{\frac{1}{NT}\sum_{i=1}^{N}\left\|W_{X,i}\right\|\left\|W_{\varepsilon,i}\right\|\right\}\left\{\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\right\}\frac{1}{T^{\frac{1}{2}}}\left\|\tilde{W}_{F^0}\right\|\left\|\tilde{Q}\right\|\left\|\tilde{H}^{-1}\right\|\\
&\leq O_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\left\|\tilde{Q}\right\|\left\|\tilde{H}^{-1}\right\|.
\end{aligned}$$

Then it remains to study the order of $\left\|\tilde{Q}\right\|$. To do that, we consider

$$\frac{1}{T}\tilde{W}_{F^0}^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)=O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right),\tag{B.2.8}$$

and similarly

$$\frac{1}{T}\hat{W}_F^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)=O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right).\tag{B.2.9}$$

Then

$$\begin{aligned}
&\left\|\mathbb{I}_R-\tilde{H}^{*}\frac{\tilde{W}_{F^0}^{*}\tilde{W}_{F^0}}{T}\tilde{H}\right\|=\left\|\frac{1}{T}\hat{W}_F^{*}\hat{W}_{F^0}-\tilde{H}^{*}\frac{\tilde{W}_{F^0}^{*}\tilde{W}_{F^0}}{T}\tilde{H}\right\| \qquad \text{(B.2.10)}\\
&\leq\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\right\|+\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}\tilde{W}_{F^0}\tilde{H}\right\|\\
&+\left\|\tilde{H}^{*}\tilde{W}_{F^0}^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\right\|\\
&=\left\|\tilde{H}\right\|O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)=O_p\left(\gamma_L^{1-2d_{F,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right).
\end{aligned}$$

and so is the probability order of $\left\|\tilde{Q}\right\|$. Therefore we can conclude that

$$\begin{aligned}
c_2 &= \sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}-d_\varepsilon+2d_{F,\min}-2d_{F,\max}}\delta_{W1,NT}^{2}\left\|\tilde{\delta}\right\|^{2}\right)\\
&\quad+O_p\left(\gamma_L^{d_{Z,\min}-d_{X,\max}+2d_{F,\min}-2d_{F,\max}}\sqrt{\frac{L}{N}}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon}\right)\\
&\quad=o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|\left\|\tilde{\delta}\right\|\right)+o_p\left(1\right).
\end{aligned}$$

Next for c_1 we have

c_1

$$= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \mathrm{Re}\left(\frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \left(\frac{\tilde{W}_{F^0}^* \tilde{W}_{F^0}}{T} \right)^{-1} \left(\hat{W}_F \tilde{H}^{-1} - \tilde{W}_{F^0} \right)^* W_{\varepsilon,i} \right)$$

$$= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \mathrm{Re}\left(\frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \left(\frac{\tilde{W}_{F^0}^* \tilde{W}_{F^0}}{T} \right)^{-1} \tilde{\Gamma}_F \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \tilde{\Gamma}_F \left(\frac{\hat{W}_F^* \tilde{W}_{F^0}}{T} \right)^{-1} \left(\tilde{I}_1 + \cdots + \tilde{I}_8 \right)^* W_{\varepsilon,i} \right)$$

$$\equiv c_{1,1} + \cdots + c_{1,8}.$$

Let

$$\breve{G} = \tilde{G} \left(\frac{\hat{W}_F^* \tilde{W}_{F^0}}{T} \right)^{-1} = \left(\frac{\tilde{W}_{F^0}^* \tilde{W}_{F^0}}{T} \right)^{-1} \tilde{\Gamma}_F \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \tilde{\Gamma}_F \left(\frac{\hat{W}_F^* \tilde{W}_{F^0}}{T} \right)^{-1}.$$

Then

$$\begin{aligned}
\|c_{1,1}\| &= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} I_1^* W_{\varepsilon,i} \right\| \\
&= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} \hat{W}_F \frac{1}{NT} \sum_{k=1}^{N} W_{X,k} \tilde{\delta}\tilde{\delta}' W_{X,k}^* W_{\varepsilon,i} \right\| \\
&\le \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \|\Gamma_Z\| \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|W_{X,i}\| \|W_{\varepsilon,i}\| \right\} \frac{1}{NT} \sum_{k=1}^{N} \|W_{X,k}\|^2 \frac{1}{T^{\frac{1}{2}}} \left\| \tilde{W}_{F^0} \right\| \frac{1}{T^{\frac{1}{2}}} \left\| \hat{W}_F \right\| \left\| \tilde{\delta} \right\|^2 \left\| \breve{G} \right\| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon + d_{Z,\min} - 1} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|W_{X,i}\| \|W_{\varepsilon,i}\| \right\} \frac{1}{NT} \sum_{k=1}^{N} \|W_{X,k}\|^2 \left\| \breve{G} \right\| \left\| \tilde{\delta} \right\|^2 \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon + d_{Z,\min} - 1} O_p\left(\gamma_L^{1 - d_{X,\max} - d_\varepsilon} \right) O_p\left(\gamma_L^{1 - 2d_{X,\max}} \right) O_p\left(\gamma_L^{2d_{F,\min} - 1} \right) \left\| \tilde{\delta} \right\|^2 \\
&= O_p\left(\sqrt{NL}\gamma_L^{d_{Z,\min} + 2d_{F,\min} - 3d_{X,\max}} \left\| \tilde{\delta} \right\|^2 \right) = o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right),
\end{aligned}$$

where we use the fact that $\left\| \breve{G} \right\| \lesssim \left\| \tilde{\Gamma}_F \right\|^2 = O_p\left(\gamma_L^{2d_{F,\min} - 1} \right)$. Next, for $c_{1,2}$ we have

$$\begin{aligned}
\|c_{1,2}\| &= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} I_2^* W_{\varepsilon,i} \right\| \\
&= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} \hat{W}_F^* W_{F^0} \frac{1}{NT} \sum_{k=1}^{N} \lambda_k \tilde{\delta}' W_{X,k} W_{\varepsilon,i} \right\| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \|\Gamma_Z\| \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|W_{X,i}\| \|W_{\varepsilon,i}\| \right\} \left\{ \frac{1}{NT^{\frac{1}{2}}} \sum_{k=1}^{N} \|W_{X,k}\| \|\lambda_k\| \right\} \left\{ \frac{1}{T^{\frac{1}{2}}} \|W_{F^0}\| \right\} \left\| \breve{G} \right\| \left\| \tilde{\delta} \right\| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon + d_{Z,\min} - 1} O_p\left(\gamma_L^{1 - d_{X,\max} - d_\varepsilon} \right) O_p\left(\gamma_L^{1/2 - d_{X,\max}} \right) O_p\left(\gamma_L^{\frac{1}{2} - d_{F,\max}} \right) O_p\left(\gamma_L^{2d_{F,\min} - 1} \right) \left\| \tilde{\delta} \right\| \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} O_p\left(\gamma_L^{2d_{F,\min} + 2d_{Z,\min} - d_{F,\max} - 2d_{X,\max} - d_\varepsilon} \left\| \tilde{\delta} \right\| \right) \\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right).
\end{aligned}$$

Similarly, we can show that $c_{1,3}$ to $c_{1,5}$ are each $o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right)$. For $c_{1,6}$, we have

$$\|c_{1,6}\| = \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} I_6^* W_{\varepsilon,i} \right\|$$

$$= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} \hat{W}_F^* \frac{1}{NT} \sum_{k=1}^{N} W_{\varepsilon,k} \lambda_k' W_{F^0}^* W_{\varepsilon,i} \right\|$$

$$\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \Gamma_Z \right\| \frac{1}{NT} \sum_{i=1}^{N} \left\| W_{X,i} \right\| \left\| W_{F^0}^* W_{\varepsilon,i} \right\| \left\| \frac{1}{NT^{\frac{3}{2}}} \sum_{k=1}^{N} \hat{W}_F^* W_{\varepsilon,k} \lambda_k' \right\| \left\| \breve{G} \right\|$$

$$= \sqrt{NL}\gamma_L^{d_\varepsilon + d_{Z,\min} - 1} O_p\left(L^{-\frac{1}{2}} \gamma_L^{1 - d_{X,\max} - d_\varepsilon}\right)$$

$$\times O_p\left(\delta_{W,NL} \left\| \tilde{\delta} \right\| + N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2} - 2d_{F,\max} - d_\varepsilon}\right) O_p\left(\gamma_L^{d_{F,\min} - d_{F,\max}}\right)$$

$$= \sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} O_p\left(L^{-\frac{1}{2}} \gamma_L^{2d_{Z,\min} - d_{X,\max} - d_\varepsilon + d_{F,\min} - d_{F,\max}} \delta_{W,NL} \left\| \tilde{\delta} \right\|\right)$$

$$+ O_p\left(L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2} - 2d_{F,\max} - d_\varepsilon + d_{Z,\min} - d_{X,\max} + d_{F,\min} - d_{F,\max}}\right)$$

$$= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\|\right) + o_p(1)$$

by Lemma B.1.7 where $\delta_{W,NL} = N^{-\frac{1}{2}} \gamma_L^{1 - d_{X,\max} - d_{F,\max} - d_\varepsilon} \left(\gamma_L^{\frac{1}{2} - d_{F,\max}} + \gamma_L^{\frac{1}{2} - d_\varepsilon}\right)$. For $c_{1,7}$, with a nonrandom P-vector ω such that $\|\omega\| = 1$, we have

$$|\omega' c_{1,7}|$$

$$= \left| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1} \omega' \Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} I_7^* W_{\varepsilon,i} \right|$$

$$= \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \frac{1}{NT} \left| \mathrm{tr}\left(\tilde{W}_{F^0} \left(\frac{\tilde{W}_{F^0}^* \tilde{W}_{F^0}}{T} \right)^{-1} \tilde{\Gamma}_F \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \lambda_k W_{\varepsilon,k}^* \sum_{i=1}^{N} \frac{W_{\varepsilon,i} \omega' \Gamma_Z W_{X,i}^*}{T} \right) \right|$$

$$\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \tilde{\Gamma}_F \right\| \frac{1}{NT} \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k W_{\varepsilon,k}^* \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{W_{\varepsilon,i} \omega' \Gamma_Z W_{X,i}^* \tilde{W}_{F^0}}{T} \right\|$$

$$\leq \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \tilde{\Gamma}_F \right\|}{NT} \sum_{l=1}^{L} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} A_i W_{\varepsilon,il} \right\| \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k W_{\varepsilon,kl}^* \right\|$$

$$\leq \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \tilde{\Gamma}_F \right\|}{N} \left(\frac{1}{T} \sum_{l=1}^{L} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} A_i W_{\varepsilon,il} \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{l=1}^{L} \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k W_{\varepsilon,kl}^* \right\|^2 \right)^{\frac{1}{2}}$$

$$= O_p\left(\sqrt{\frac{L}{N}} \gamma_L^{\frac{1}{2} + d_{Z,\min} + d_{F,\min} - d_{X,\max} - 2d_\varepsilon} \right) = o_p(1),$$

where $A_i = \frac{\omega' \Gamma_Z W_{X,i}^* \tilde{W}_{F^0}}{T}$. Note that this term corresponds to one of the asymptotic bias in the time domain LS estimator but it asymptotically negligible here due to the smaller order of magnitude for L. To make our asymptotic theory more comparable with the one in time domain, we keep this term explicit. The last two equalities hold by the following reasoning:

$$E\left(\frac{1}{T} \sum_{l=1}^{L} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} A_i W_{\varepsilon,il} \right\|^2 \right) \leq \frac{1}{NT} \sum_{l=1}^{L} \sum_{i,k=1}^{N} \left| E\left(A_i A_k^*\right) \right| \left| E\left(W_{\varepsilon,il} W_{\varepsilon,kl}^*\right) \right|,$$

$$\leq \frac{1}{NT} \sum_{l=1}^{L} \sum_{i,k=1}^{N} E\left|A_i A_k^*\right| \left(E\left| W_{\varepsilon,il} W_{\varepsilon,kl}^* \right|^2 \right)^{\frac{1}{2}} \quad \text{(B.2.11)}$$

where $E\left|W_{\varepsilon,il}W^*_{\varepsilon,kl}\right|^2 \leq \gamma_l^{-4d_\varepsilon}\left(\overline{\sigma}^W_{ik}\right)^2$ by Assumption C*(ii). For $E\left|A_iA_k^*\right|$ we also have

$$E\left|A_iA_k^*\right| \leq E\left(\|A_i\|\,\|A_k\|\right) \leq \sqrt{E\|A_i\|^2 E\|A_k\|^2}.$$

Focusing on $E\|A_i\|^2$, we denote $\overline{W}_{X,ij}$ as the conjugate of $W_{X,ij}$, and denote $\check{W}_{F^0,j} = \check{\Gamma}_{F,j}\check{W}_{F^0,j}$ as in Assumption C*(iii) with $\check{W}_{X,ij}$ defined in the same manner. Then

$$\begin{aligned}
E\|A_i\|^2 &= \frac{1}{T^2}\omega'\Gamma_Z\sum_{j,l=1}^{L} E\left(\overline{W}_{X,ij}W'_{F^0,j}\tilde{\Gamma}_F^2 W_{F^0,l}W^*_{X,il}\right)\Gamma_Z\omega \\
&= \frac{1}{T^2}\omega'\Gamma_Z\sum_{j,l=1}^{L} E\left(\check{\Gamma}^{-1}_{X,j}\check{\overline{W}}_{X,ij}\check{W}'_{F^0,j}\check{\Gamma}^{-1}_{F,j}\tilde{\Gamma}^2_F\check{\Gamma}^{-1}_{F,l}\check{W}_{F^0,l}\check{W}^*_{X,il}\check{\Gamma}^{-1}_{X,l}\right)\Gamma_Z\omega \\
&\lesssim \frac{1}{T^2}\gamma_L^{2d_{Z,\min}}\sum_{j,l=1}^{L}\gamma_j^{-d_{X,\max}}\gamma_l^{-d_{X,\max}}\left\|\check{\Gamma}^{-1}_{F,j}\tilde{\Gamma}^2_F\check{\Gamma}^{-1}_{F,l}\right\| E\left(\left\|\check{\overline{W}}_{X,ij}\check{W}'_{F^0,j}\right\|\left\|\check{W}_{F^0,l}\check{W}^*_{X,il}\right\|\right) \\
&= O\left(\gamma_L^{2-2d_{X,\max}+2d_{Z,\min}}\right)
\end{aligned}$$

by Assumption B*(i) and B*(iii). Therefore what remains is to consider the order of

$$\frac{1}{NT}\sum_{l=1}^{L}\sum_{i,k=1}^{N}\left(E\left|W_{\varepsilon,il}W^*_{\varepsilon,kl}\right|^2\right)^{\frac{1}{2}},$$

which is given by $O\left(\gamma_L^{1-2d_\varepsilon}\right)$. This implies that (B.2.11) is $O\left(\gamma_L^{3-2d_{X,\max}-2d_\varepsilon+2d_{Z,\min}}\right)$. Similarly, we have

$$E\left(\frac{1}{T}\sum_{l=1}^{L}\left\|\frac{1}{\sqrt{N}}\sum_{i=k}^{N}\lambda_k W_{\varepsilon,kl}\right\|^2\right) = O\left(\gamma_L^{1-2d_\varepsilon}\right),$$

which altogether forms the order of $c_{1,7}$. Then lastly $c_{1,8}$ is given by

$$\begin{aligned}
c_{1,8} &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W^*_{X,i}\tilde{W}_{F^0}}{T}\check{G}I_8^*W_{\varepsilon,i} \\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W^*_{X,i}\tilde{W}_{F^0}}{T}\check{G}\hat{W}_F^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W^*_{\varepsilon,k}W_{\varepsilon,i} \\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W^*_{X,i}\tilde{W}_{F^0}}{T}\check{G}\tilde{H}^*\tilde{W}^*_{F^0}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W^*_{\varepsilon,k}W_{\varepsilon,i} \\
&+ \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W^*_{X,i}\tilde{W}_{F^0}}{T}\check{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W^*_{\varepsilon,k}W_{\varepsilon,i} \equiv c_{1,81}+c_{1,82}.
\end{aligned}$$

Then it remains to study $c_{1,81}$ and $c_{1,82}$. For $c_{1,81}$, we have

$$c_{1,81} = \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W^*_{X,i}\tilde{W}_{F^0}}{T}\check{G}\tilde{H}^*\tilde{W}^*_{F^0}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}\left[W^*_{\varepsilon,k}W_{\varepsilon,i}-E\left(W^*_{\varepsilon,k}W_{\varepsilon,i}\right)\right]$$

$$+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\tilde{H}^*\tilde{W}_{F^0}^*\frac{1}{NT}\sum_{k=1}^N W_{\varepsilon,k}E\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)\equiv c_{1,811}+c_{1,812}.$$

Note that

$$\begin{aligned}
&\|c_{1,811}\|\\
&=\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{k=1}^N\left[\frac{1}{N\sqrt{T}}\sum_{i=1}^N\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\sum_{l=1}^L\left[W_{\varepsilon,kl}^*W_{\varepsilon,il}-E\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)\right]\right]\breve{G}\tilde{H}^*\left[\frac{1}{\sqrt{T}}\tilde{W}_{F^0}^*W_{\varepsilon,k}\right]\right\|\\
&\lesssim\sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|\left(\frac{1}{NT}\sum_{k=1}^N\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\sum_{l=1}^L\left[W_{\varepsilon,kl}^*W_{\varepsilon,il}-E\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)\right]\right\|^2\right)^{\frac{1}{2}}\\
&\times\left(\frac{1}{NT}\sum_{k=1}^N\left\|\frac{1}{\sqrt{T}}\tilde{W}_{F^0}^*W_{\varepsilon,k}\right\|^2\right)^{\frac{1}{2}}\\
&=O_p\left(\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\right)O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_\varepsilon-d_{X,\max}}\right)O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)\\
&=O_p\left(\frac{1}{\sqrt{T}}\gamma_L^{1-2d_\varepsilon+d_{Z,\min}-d_{X,\max}}\right)=o_p(1)
\end{aligned}$$

by Assumption D*(iii), where the last two equalities hold by Assumption C*(i) and the fact that

$$\begin{aligned}
&E\left(\frac{1}{NT}\sum_{k=1}^N\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\sum_{l=1}^L\left[W_{\varepsilon,kl}^*W_{\varepsilon,il}-E\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)\right]\right\|^2\right)\\
&=O\left(N^{-1}\gamma_L^{2-4d_\varepsilon}\gamma_L^{1-2d_{X,\max}}\right) \qquad \text{(B.2.12)}
\end{aligned}$$

by Assumption C*(v) following the similar reasoning to (B.2.3) in the proof of Lemma B.1.4. Similarly $c_{1,812}$ has the same order, which is obtained by replacing $W_{\varepsilon,kl}^*W_{\varepsilon,il}-E\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)$ on the left hand side of (B.2.12) by $E\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)$ and using Assumption C*(ii). And for $c_{1,82}$, we have

$$\begin{aligned}
c_{1,82}&=\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\frac{1}{NT}\sum_{k=1}^N W_{\varepsilon,k}\left[W_{\varepsilon,k}^*W_{\varepsilon,i}-E\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)\right]\\
&+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\frac{1}{NT}\sum_{k=1}^N W_{\varepsilon,k}E\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)\\
&\equiv c_{1,821}+c_{1,822},
\end{aligned}$$

By Assumption B*(iii) and B*(iv), we have

$$\|c_{1,821}\|=\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\frac{1}{NT}\sum_{k=1}^N W_{\varepsilon,k}\left[W_{\varepsilon,k}^*W_{\varepsilon,i}-E\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)\right]\right\|$$

$$\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|\gamma_L^{2d_{F,\min}-1}\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\sum_{l=1}^{L}\left[W_{\varepsilon,kl}^*W_{\varepsilon,il}-E\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)\right]\right\|^2\right)^{\frac{1}{2}}$$

$$\times\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,k}\right\|^2\right)^{\frac{1}{2}}$$

$$=\sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|O_p\left(\gamma_L^{2d_{F,\min}-1}\right)O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_\varepsilon-d_{X,\max}}\right)$$

$$\times O_p\left(\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$$

$$=O_p\left(\alpha_{3,NL}\left\|\tilde{\delta}\right\|+\sqrt{\frac{L}{N}}\gamma_L^{1-d_{X,\max}-d_{F,\max}+d_{Z,\min}+2d_{F,\min}-3d_\varepsilon}\right)$$

$$=o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|\left\|\tilde{\delta}\right\|\right)+o_p(1),$$

by Lemma B.1.6, as

$$E\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,k}\right\|^2\right)=O\left(\left(\delta_{W1,NT}^2\left\|\tilde{\delta}\right\|^2+N^{-1}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon}\right)\gamma_L^{1-2d_\varepsilon}\right),$$

and $\alpha_{3,NL}=\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|N^{-\frac{1}{2}}\gamma_L^{2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_\varepsilon}$. Next for $c_{1,822}$, we have by the same reasoning and conditions that

$$\|c_{1,822}\|$$

$$=\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}E\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)\right\|$$

$$\lesssim\sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|\gamma_L^{2d_{F,\min}-1}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\right\|\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,k}\right\|\left|\frac{E\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)}{T}\right|$$

$$\leq\sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|\gamma_L^{2d_{F,\min}-1}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,k}\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\right\|^2\right\}^{1/2}$$

$$\times\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{E\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)}{T}\right\|^2\right\}^{1/2}$$

$$=\sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|\gamma_L^{2d_{F,\min}-1}O_p\left(\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$$

$$\times O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right)O_p\left(N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right)$$

$$=\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}O_p\left(N^{-\frac{1}{2}}\gamma_L^{2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_\varepsilon}\delta_{W1,NT}\left\|\tilde{\delta}\right\|\right)$$

$$+O_p\left(\sqrt{\frac{L}{N}}\gamma_L^{1+d_{Z,\min}+2d_{F,\min}-d_{X,\max}-d_{F,\max}-3d_\varepsilon}\right)$$

$$=o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|\left\|\tilde{\delta}\right\|\right)+o_p(1)$$

by Lemma B.1.6 and by Assumption C*(ii).

Lastly for d we have

$$\begin{aligned}
\|d\| &= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T} \tilde{Q} \tilde{W}_{F^0}^{*} W_{\varepsilon,i} \right\| \\
&\leq \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \|\Gamma_Z\| \left\| \tilde{Q} \right\| \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T} \right\| \left\| \frac{\tilde{W}_{F^0}^{*} W_{\varepsilon,i}}{T} \right\| \\
&\leq \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \|\Gamma_Z\| \left\| \tilde{Q} \right\| \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T} \right\|^2 \right\}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\tilde{W}_{F^0}^{*} W_{\varepsilon,i}}{T} \right\|^2 \right\}^{1/2} \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \|\Gamma_Z\| O_p\left(\gamma_L^{1-2d_{F,\max}} \left(\delta_{W1,NT} \left\| \tilde{\delta} \right\| + N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon} \right) \right) \\
&\times O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}} \right) O_p\left(L^{-\frac{1}{2}} \gamma_L^{\frac{1}{2}-d_\varepsilon} \right) \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} O_p\left(L^{-\frac{1}{2}} \gamma_L^{1-d_{X,\max}-d_\varepsilon+2d_{Z,\min}-2d_{F,\max}} \delta_{W1,NT} \left\| \tilde{\delta} \right\| \right) \\
&+ O_p\left(\gamma_L^{2-d_{X,\max}-d_{F,\max}-2d_\varepsilon+d_{Z,\min}-d_{F,\max}} \right) \\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right) + o_p(1),
\end{aligned}$$

by Assumption D*(iii) as before.

This completes the proof of approximation for the first part as

$$\begin{aligned}
&\frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^{*} \mathbf{M}_{\hat{W}_F} W_{\varepsilon,i} \right) \\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^{*} \mathbf{M}_{W_{F^0}} W_{\varepsilon,i} \right) + o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right) + o_p(1).
\end{aligned}$$

Then for the second part given by

$$\begin{aligned}
&\frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left[\left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} W_{X,k}^{*} \right) \left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{W_{F^0}} \right) W_{\varepsilon,i} \right] \\
&\equiv \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left[W_{V,i}^{*} \left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{W_{F^0}} \right) W_{\varepsilon,i} \right].
\end{aligned}$$

By replacing $W_{X,i}$ by $W_{V,i}$, we can obtain the same order for the second part. Then we can conclude that

$$\begin{aligned}
&\frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(\left(W_{X,i}^{*} \mathbf{M}_{\hat{W}_F} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} W_{X,k}^{*} \mathbf{M}_{\hat{W}_F} \right) W_{\varepsilon,i} \right) \\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re}\left(\left(W_{X,i}^{*} \mathbf{M}_{W_{F^0}} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} W_{X,k}^{*} \mathbf{M}_{W_{F^0}} \right) W_{\varepsilon,i} \right) \\
&+ o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} \left\| \tilde{\delta} \right\| \right) + o_p(1).
\end{aligned}$$

This completes the proof of Lemma B.1.9. ■

B.3 Demeaned Time Domain Least Squares Estimator

By (3.2.2) and (3.2.3), the model (3.2.1) can be rewritten as

$$\begin{aligned} Y_{it} &= X_{it}'\beta^0 + \lambda_i' F_t^0 + \varepsilon_{it} \\ &= \left(\mu_{X,i} + \widetilde{X}_{it}\right)'\beta^0 + \lambda_i'\left(\mu_F + \widetilde{F}_t^0\right) + \varepsilon_{it} \\ &= \widetilde{X}_{it}'\beta^0 + \lambda_i'\widetilde{F}_t^0 + \left(\mu_{X,i}'\beta^0 + \lambda_i'\mu_F\right) + \varepsilon_{it} \\ &\equiv \widetilde{X}_{it}'\beta^0 + \lambda_i'\widetilde{F}_t^0 + \tilde{\mu}_i + \varepsilon_{it}, \end{aligned} \tag{B.3.1}$$

which is a factor model together with additive individual effects. Then following Bai (2009), we conduct the LS estimation to its demeaned version

$$\dot{Y}_{it} = \dot{\widetilde{X}}_{it}'\beta^0 + \lambda_i'\dot{\widetilde{F}}_t^0 + \dot{\varepsilon_{it}}, \tag{B.3.2}$$

where $\dot{Y}_{it} = Y_{it} - \overline{Y}_{i\cdot}$ and $\overline{Y}_{i\cdot} = \frac{1}{T}\sum_{t=1}^T Y_{it}$. In this study both $\tilde{\mu}_i$ and $\widetilde{F}_t^0$ are nuisance parameters and thus we do not need the identification condition $\sum_{t=1}^T \widetilde{F}_t^0 = 0$, which is quite restrictive in application. And as in (3.2.5) and (3.2.6), the LS estimator of model (B.3.2) is given by

$$\beta^* = \left(\sum_{i=1}^N \dot{\widetilde{X}}_i'\mathbf{M}_{F^*}\dot{\widetilde{X}}_i\right)^{-1}\sum_{i=1}^N \dot{\widetilde{X}}_i'\mathbf{M}_{F^*}\dot{Y}_i$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^N\left(\dot{Y}_i - \dot{\widetilde{X}}_i\beta^*\right)\left(\dot{Y}_i - \dot{\widetilde{X}}_i\beta^*\right)'\right]F^* = F^*V_{NT}.$$

In the following theorem we present the asymptotic behavior of β^*. As before, we denote

$$Z_i^* = \mathbf{M}_{\dot{\widetilde{F}}^0}\dot{\widetilde{X}}_i - \frac{1}{N}\sum_{k=1}^N a_{ik}\mathbf{M}_{\dot{\widetilde{F}}^0}\dot{\widetilde{X}}_k.$$

And the theorem is stated as below.

Theorem B.3.1 *Suppose that Assumption A-F hold. Then for comparable N and T such that $T/N \to \rho > 0$, we have for some positive definite matrices $\tilde{D}_0$ and $\tilde{\Sigma}$ that,*

(i) when $d_Z + d_\varepsilon > \frac{1}{2}$ and $d_F + d_\varepsilon > \frac{1}{2}$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T^{1-(d_Z+d_\varepsilon)}T^{1-(d_F+d_\varepsilon)}}C_1^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where B^* is the probability limit of

$$\tilde{B}^* = -D\left(\dot{\tilde{F}}^0\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} \frac{\left(\dot{\tilde{X}}_i - \dot{\tilde{V}}_i\right)' \dot{\tilde{F}}^0}{T} \left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1} \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{T}\sum_{k=1}^{N} \lambda_k \dot{\varepsilon_k}' \dot{\varepsilon}_i$$

with $\dot{\tilde{V}}_i = \frac{1}{N}\sum_{k=1}^{N} a_{ik}\dot{\tilde{X}}_k$; and C_1^* is the probability limit of

$$\tilde{C}_1^* = -D\left(\dot{\tilde{F}}^0\right)^{-1} \frac{1}{NT^{d_Z+d_\varepsilon}} \sum_{i=1}^{N} \dot{\tilde{X}}_i' \mathbf{M}_{\dot{\tilde{F}}^0} \frac{1}{NT^{d_F+d_\varepsilon}} \sum_{k=1}^{N} \dot{\varepsilon_k}\dot{\varepsilon_k}' \dot{\tilde{F}}^0 \left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1} \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \lambda_i;$$

(ii) when $d_Z + d_\varepsilon > \frac{1}{2} \geq d_F + d_\varepsilon$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T^{1-(d_Z+d_\varepsilon)}T^{\frac{1}{2}}}C_2^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where B^* is the same as above and C_2^* is the probability limit of

$$\tilde{C}_2^* = -D\left(\dot{\tilde{F}}^0\right)^{-1} \frac{1}{NT^{d_Z+d_\varepsilon}} \sum_{i=1}^{N} \dot{\tilde{X}}_i' \mathbf{M}_{\dot{\tilde{F}}^0} \frac{1}{NT^{\frac{1}{2}}} \sum_{k=1}^{N} \dot{\varepsilon_k}\dot{\varepsilon_k}' \dot{\tilde{F}}^0 \left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1} \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \lambda_i;$$

(iii) when $d_F + d_\varepsilon > \frac{1}{2} \geq d_Z + d_\varepsilon$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T^{1-(d_F+d_\varepsilon)}T^{\frac{1}{2}}}C_3^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where B^* is the same as above and C_3^* is the probability limit of

$$\tilde{C}_3^* = -D\left(\dot{\tilde{F}}^0\right)^{-1} \frac{1}{NT^{\frac{1}{2}}} \sum_{i=1}^{N} \dot{\tilde{X}}_i' \mathbf{M}_{\dot{\tilde{F}}^0} \frac{1}{NT^{d_F+d_\varepsilon}} \sum_{k=1}^{N} \dot{\varepsilon_k}\dot{\varepsilon_k}' \dot{\tilde{F}}^0 \left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1} \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \lambda_i;$$

(iv) when $d_Z + d_\varepsilon \leq \frac{1}{2}$ and $d_F + d_\varepsilon \leq \frac{1}{2}$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T}C_4^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where B^* is the same as above and C_4^* is the probability limit of

$$\tilde{C}_4^* = -D\left(\dot{\tilde{F}}^0\right)^{-1} \frac{1}{NT^{\frac{1}{2}}} \sum_{i=1}^{N} \dot{\tilde{X}}_i' \mathbf{M}_{\dot{\tilde{F}}^0} \frac{1}{NT^{\frac{1}{2}}} \sum_{k=1}^{N} \dot{\varepsilon_k}\dot{\varepsilon_k}' \dot{\tilde{F}}^0 \left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1} \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \lambda_i.$$

By within-group transformation, the LS estimator now obtains a unified convergence rate that only depends on the memory parameters of idiosyncratic error term. But meanwhile the order of bias term has a complex pattern as it depends both on d_Z+d_ε and d_F+d_ε. Therefore to sum up the LS estimator using data in time domain could be difficult to implement if we account for long memory in the model.

Proof of Theorem B.3.1. The proof of Theorem B.3.1 can follow the same steps as the proof of Theorem 3.3.1. It is easy to see that all the asymptotically negligible terms are still negligible, and thus we can focus on the order of bias terms and the convergence rate of β^*. To be specific, the order of two bias terms under LS estimator of model (B.3.2), B^* and C_j^*, $j=1,\dots,4$, are related to the order of the following two terms

$$\widetilde{c1.7}^* = -\frac{1}{NT}\sum_{i=1}^{N}\frac{\left(\dot{\tilde{X}}_i-\dot{\tilde{V}}_i\right)'\dot{\tilde{F}}^0}{T}\left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{N}\sum_{k=1}^{N}\lambda_k\dot{\varepsilon}_k{}'\dot{\varepsilon}_i, \tag{B.3.3}$$

and

$$\widetilde{J8}^* = -\frac{1}{NT}\sum_{i=1}^{N}\dot{\tilde{X}}_i'\mathbf{M}_{\dot{\tilde{F}}^0}\frac{1}{NT}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k{}'\dot{\tilde{F}}^0\left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i \tag{B.3.4}$$

respectively. Firstly for (B.3.3), the same arguments can hold as in the non-demean model, and thus $\widetilde{c1.7}^* = O_p\left(\frac{1}{N}\right)$. And then for (B.3.4), we firstly denote $\dot{\tilde{Z}}_i' = \dot{\tilde{X}}_i'\mathbf{M}_{\dot{\tilde{F}}^0}$. Then by definition of Z_i^*, we can see the memory parameter vector of $\dot{\tilde{Z}}_i$ is also d_Z, which implies that

$$\widetilde{J8}^* = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\dot{\tilde{Z}}_{it}\dot{\varepsilon}_{kt}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\dot{\tilde{F}}_t^0\dot{\varepsilon}_{kt}\right)\left(\frac{\dot{\tilde{F}}^{0\prime}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,$$

where $\frac{1}{T}\sum_{t=1}^{T}\dot{\tilde{Z}}_{it}\dot{\varepsilon}_{kt}$ and $\frac{1}{T}\sum_{t=1}^{T}\dot{\tilde{F}}_t^0\dot{\varepsilon}_{kt}$ can be treated as the sample cross-covariance between $\dot{\tilde{Z}}_{it}$, $\dot{\tilde{F}}_t^0$ and ε_{kt}. Therefore Assumption D(i) implies that

$$\frac{1}{T}\sum_{t=1}^{T}\dot{\tilde{F}}_t^0\dot{\varepsilon}_{kt} = \begin{cases} O_p\left(T^{d_F+d_\varepsilon-1}\right), \text{ if } d_F+d_\varepsilon>\frac{1}{2} \\ O_p(T^{-\frac{1}{2}}), \text{ if } d_F+d_\varepsilon\le\frac{1}{2}\end{cases},$$

which concludes that

$$\widetilde{J8}^* = \begin{cases} O_p\left(T^{d_Z+d_\varepsilon-1}T^{d_F+d_\varepsilon-1}\right), \text{ if } d_Z+d_\varepsilon>\frac{1}{2} \text{ and } d_F+d_\varepsilon>\frac{1}{2} \\ O_p\left(T^{d_Z+d_\varepsilon-1}T^{-\frac{1}{2}}\right), \text{ if } d_Z+d_\varepsilon>\frac{1}{2}\ge d_F+d_\varepsilon \\ O_p\left(T^{d_F+d_\varepsilon-1}T^{-\frac{1}{2}}\right), \text{ if } d_F+d_\varepsilon>\frac{1}{2}\ge d_Z+d_\varepsilon \\ O_p(T^{-1}), \text{ if } d_Z+d_\varepsilon\le\frac{1}{2} \text{ and } d_F+d_\varepsilon\le\frac{1}{2}\end{cases}.$$

Therefore by what we have so far, the asymptotic representation of $\beta^* - \beta^0$ follows that

$$\beta^* - \beta^0 = D\left(\dot{\tilde{F}}^0\right)^{-1}\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i^{*'}\dot{\varepsilon}_i + \widetilde{c1.7}^* + \widetilde{J8}^*\right] + o_p(1) + o_p\left(\beta^* - \beta^0\right),$$

and Assumption F implies that $N^{-\frac{1}{2}}T^{d_\varepsilon - \frac{1}{2}}\sum_{i=1}^{N} Z_i^{*'}\dot{\varepsilon}_i \xrightarrow{d} \mathcal{N}\left(0, \tilde{\Sigma}\right)$. And also $D\left(\dot{\tilde{F}}^0\right) \xrightarrow{p} \tilde{D}_0$ for some positive definite matrix $\tilde{D}_0$. Then we have the asymptotic distribution of $\beta^* - \beta^0$ given by

$$\begin{cases} N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T^{1-(d_Z+d_\varepsilon)}T^{1-(d_F+d_\varepsilon)}}C_1^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right), \text{ if } \min(d_Z, d_F) + d_\varepsilon > \frac{1}{2} \\ N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T^{1-(d_Z+d_\varepsilon)}T^{\frac{1}{2}}}C_2^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right), \text{ if } d_Z + d_\varepsilon > \frac{1}{2} \geq d_F + d_\varepsilon \\ N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T^{1-(d_F+d_\varepsilon)}T^{\frac{1}{2}}}C_3^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right), \text{ if } d_F + d_\varepsilon > \frac{1}{2} \geq d_Z + d_\varepsilon \\ N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^* - \beta^0 - \frac{1}{N}B^* - \frac{1}{T}C_4^*\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right), \text{ if } d_Z + d_\varepsilon \leq \frac{1}{2} \text{ and } d_F + d_\varepsilon \leq \frac{1}{2} \end{cases}$$

where B^* is the probability limit of

$$\tilde{B}^* = -D\left(\dot{\tilde{F}}^0\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\dot{\tilde{X}}_i - \dot{\tilde{V}}_i\right)'\dot{\tilde{F}}^0}{T}\left(\frac{\dot{\tilde{F}}^{0'}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{T}\sum_{k=1}^{N}\lambda_k\dot{\varepsilon}_k{}'\dot{\varepsilon}_i,$$

and C_1^* is the probability limit of

$$\tilde{C}_1^* = -D\left(\dot{\tilde{F}}^0\right)^{-1}\frac{1}{NT^{d_Z+d_\varepsilon}}\sum_{i=1}^{N}\dot{\tilde{X}}_i{}'\mathbf{M}_{\dot{\tilde{F}}^0}\frac{1}{NT^{d_F+d_\varepsilon}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k{}'\dot{\tilde{F}}^0\left(\frac{\dot{\tilde{F}}^{0'}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,$$

and C_2^* is the probability limit of

$$\tilde{C}_2^* = -D\left(\dot{\tilde{F}}^0\right)^{-1}\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\dot{\tilde{X}}_i{}'\mathbf{M}_{\dot{\tilde{F}}^0}\frac{1}{NT^{d_F+d_\varepsilon}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k{}'\dot{\tilde{F}}^0\left(\frac{\dot{\tilde{F}}^{0'}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,$$

C_3^* is the probability limit of

$$\tilde{C}_3^* = -D\left(\dot{\tilde{F}}^0\right)^{-1}\frac{1}{NT^{d_Z+d_\varepsilon}}\sum_{i=1}^{N}\dot{\tilde{X}}_i{}'\mathbf{M}_{\dot{\tilde{F}}^0}\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k{}'\dot{\tilde{F}}^0\left(\frac{\dot{\tilde{F}}^{0'}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,$$

and C_4^* is the probability limit of

$$\tilde{C}_4^* = -D\left(\dot{\tilde{F}}^0\right)^{-1}\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\dot{\tilde{X}}_i{}'\mathbf{M}_{\dot{\tilde{F}}^0}\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k{}'\dot{\tilde{F}}^0\left(\frac{\dot{\tilde{F}}^{0'}\dot{\tilde{F}}^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i.$$

This completes the proof of Theorem B.3.1. ■

Appendix C

Technical Results for Chapter 4

C.1 Proofs of the Main results

In this section we prove the theorems in the main text, where some auxiliary lemmas are included. Lemma C.1.1 is directly borrowed from the Lemma A.6 in the last chapter and thus presented without proof. The proof of rest of the lemmas will be given in the supplemental materials.

Lemma C.1.1 *Suppose Assumption A-F hold and the other conditions of Theorem 4.3.1 hold. Let* $\tilde{H} = \left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\left(\frac{\tilde{W}_F^*\hat{W}_F}{T}\right)V_{NL}^{-1}$*. Then*

$$T^{-\frac{1}{2}}\left\|\hat{W} - \tilde{W}_F\tilde{H}\right\| = O_p\left(\delta_{W1,NT}\left\|\tilde{\beta} - \beta\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right),$$

where $\delta_{W1,NT} = \gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$*. Furthermore, we define a new variable* Z_{it} *through its DFT by (4.3.7) and denote its memory parameter vector as* d_Z

$$\tilde{\beta} - \beta = O_p\left(\frac{T^{2d_\varepsilon}}{L}\gamma_L^{2d_{Z,\min}+d_{F,\min}-d_{X,\max}} + \frac{\gamma_L^{d_{Z,\min}-d_\varepsilon}}{\sqrt{NL}}\right).$$

Lemma C.1.2 *Suppose Assumption A-F and the other conditions of Theorem 4.3.1 hold, and define*

$$\tilde{H} = \left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\left(\frac{\tilde{W}_F\hat{W}_F^*}{T}\right)V_{NL}^{-1},$$

then $\tilde{H} = I + O_p(\delta_{NT})$ *where* $\delta_{NT} = \delta_{W1,NT}\left(\tilde{\beta} - \beta\right) + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}$ *with* $\delta_{W1,NT}$ *given by Lemma C.1.1.*

Lemma C.1.3 *Suppose Assumption A-F hold, we have for arbitrary* $1 \leq j \leq L$ *that*

$$\left\|\hat{I}_{F,j} - \tilde{I}_{F,j}\right\| \leq O_p\left(\delta_{NT}\right)\left\|\tilde{I}_{F,j}\right\| + O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right)\tilde{J}_{FF,j}$$
$$+ O_p\left(\gamma_L^{\frac{3}{2}-3d_{F,\max}}\left\|\tilde{\delta}\right\|\right)\tilde{J}_{FX,j} + O_p\left(L^{-\frac{1}{2}}\gamma_L^{\frac{5}{2}-4d_{F,\max}-d_\varepsilon}\right)\tilde{J}_{F\varepsilon,j}$$
$$+ O_p\left(\gamma_L^{\frac{3}{2}-3d_{F,\max}}\right)\check{J}_{F\varepsilon,j} + I_{E_F,j}$$

where $\tilde{J}_{FF,j} = \left\|\tilde{W}_{F,j}W^*_{F,j}\right\|$, $\tilde{J}_{FX,j} = \left(\frac{1}{N}\sum_{i=1}^N\left\|\tilde{W}_{F,j}W^*_{X,ij}\right\|^2\right)^{\frac{1}{2}}$, $\tilde{J}_{F\varepsilon,j} = \left(\frac{1}{N}\sum_{i=1}^N\left\|\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\right\|^2\right)^{\frac{1}{2}}$ *and* $\check{J}_{F\varepsilon,j} = \left\|\frac{1}{N}\sum_{i=1}^N\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\lambda_i'\right\|$; *and* $I_{E_F,j}$ *is a smaller order term at each* j *relative to the others.*

Proof of Theorem 4.3.1

Proof of this theorem is based on the proof of Theorem 1 in Robinson (1995c). Since our local Whittle estimator is conducted marginally at each $r = 1, \ldots, K$, and the proof of consistency of every $\hat{d}_{\hat{F}_r}$ is identical, we proceed with the proof in the following marginally at each $r = 1, \ldots, R$ and omit the subscript r in all the arguments for ease of notation. To be specific, for an arbitrary $\delta > 0$, let $N_\delta = \{d : |d - d_F| < \delta\}$ and $\bar{N}_\delta = (-\infty, \infty) - N_\delta$. Then

$$P\left(\left|\hat{d}_{\hat{F}} - d_F\right| \geq \delta\right) = P\left(\inf_{\bar{N}_\delta\cap\left[0,\frac{1}{2}\right)} K(d) \leq \inf_{N_\delta\cap\left[0,\frac{1}{2}\right)} K(d)\right)$$
$$\leq P\left(\inf_{\bar{N}_\delta\cap\left[0,\frac{1}{2}\right)} K(d) - K(d_F) \leq 0\right). \tag{C.1.1}$$

The consistency of $\hat{d}_{\hat{F}}$ will hold if $P\left(\left|\hat{d}_{\hat{F}} - d_F\right| \geq \delta\right)$ is $o(1)$. And by the fact that both $\hat{d}_{\hat{F}}$ and d_F lie within the interval $\left[0, \frac{1}{2}\right)$, it is sufficient to consider $\delta < \frac{1}{2}$. Next, define

$$V(d) = \log\left\{\frac{\hat{G}(d_F)}{\upsilon_F}\right\} - \log\left\{\frac{\hat{G}(d)}{G(d)}\right\} - \log\left\{\frac{1}{L}\sum_{j=1}^L j^{2(d-d_F)} \Big/ \frac{L^{2(d-d_F)}}{2(d-d_F)+1}\right\}$$
$$+ 2(d-d_F)\left\{\frac{1}{L}\sum_{j=1}^L \log j - (\log L - 1)\right\}$$
$$\equiv \log\left\{\frac{\hat{G}(d_F)}{\upsilon_F}\right\} - \log\left\{\frac{\hat{G}(d)}{G(d)}\right\} + V^*(d)$$

and

$$U(d) = 2(d-d_F) - \log\{2(d-d_F)+1\},$$

where $G(d) = \upsilon_F \frac{1}{L} \sum_{j=1}^{L} \gamma_j^{2(d-d_F)}$. Note that we focus on the profile likelihood function

$$K(d) = \log \hat{G}(d) - 2d \frac{1}{L} \sum_{j=1}^{L} \log \gamma_j$$

where $\hat{G}(d) = \frac{1}{L} \sum_{j=1}^{L} \gamma_j^{2d} \hat{I}_{F,j}$. And our profile likelihood function is different from the one in Robinson (1995c) only by replacing $I_{F,j}$, the periodogram of the true factor, with its estimation $\hat{I}_{F,j}$. Therefore the reasoning in the proof of Theorem 1 in Robinson (1995c) will continue to follow as $K(d) - K(d_F) = U(d) - V(d)$. Furthermore, we define $\tilde{\upsilon}_F = \gamma_L^{2d_F-1} \upsilon_F$ and $\tilde{G}(d) = \gamma_L^{2d_F-1} G(d)$. Then

$$\log \left\{ \frac{\hat{G}(d_F)}{\upsilon_F} \right\} - \log \left\{ \frac{\hat{G}(d)}{G(d)} \right\} = \log \left\{ \frac{\hat{G}(d_F)}{\tilde{\upsilon}_F} \right\} - \log \left\{ \frac{\hat{G}(d)}{\tilde{G}(d)} \right\},$$

which updates the expression of $V(d)$. Furthermore, (C.1.1) is bounded by

$$P\left(\inf_{\bar{N}_\delta \cap [0,\frac{1}{2})} U(d) \le \sup_{[0,\frac{1}{2})} |V(d)| \right),$$

and then it is sufficient to show that $\inf_{\bar{N}_\delta \cap [0,\frac{1}{2})} U(d)$ is lower bounded away from zero and $\sup_{[0,\frac{1}{2})} |V(d)|$ is $o(1)$. And since neither $V^*(d)$ nor $U(d)$ is random or depends on the periodogram, the orders of these two objects will also hold as indicated by Robinson (1995c) that $\inf_{\bar{N}_\delta \cap [0,\frac{1}{2})} U(d) > \frac{1}{2}\delta^2$ and $\sup_{[0,\frac{1}{2})} |V^*(d)| = o(1)$. To handle the remainder of $V(d)$, by the fact that $\tilde{\upsilon}_F = \tilde{G}(d_F)$, we will try to prove $\sup_{[0,\frac{1}{2})} \left| \log \left\{ \frac{\hat{G}(d)}{\tilde{G}(d)} \right\} \right| = o_p(1)$. Then by Taylor expansion, it is sufficient to show

$$\sup_{[0,\frac{1}{2})} \left| \frac{\hat{G}(d) - \tilde{G}(d)}{\tilde{G}(d)} \right| = o_p(1).$$

To proceed, denote $\frac{\hat{G}(d) - \tilde{G}(d)}{\tilde{G}(d)} = \frac{A(d)}{B(d)}$, where

$$A(d) = \frac{2(d-d_F)+1}{L} \sum_{j=1}^{L} \left(\frac{j}{L} \right)^{2(d-d_F)} \left(\frac{\hat{I}_{F,j}}{\tilde{g}_j} - 1 \right)$$

and

$$B(d) = \frac{2(d-d_F)+1}{L} \sum_{j=1}^{L} \left(\frac{j}{L} \right)^{2(d-d_F)},$$

as we denote $\tilde{g}_j = \tilde{\upsilon}_F \gamma_j^{-2d_F}$. Note that $B(d)$ is same as the one presented in Robinson (1995c), whose Lemma 1 then shows that $\inf_{[0,\frac{1}{2})} B(d) \ge \frac{1}{2}$. Therefore it remains to show the asymptotic negligibility of $\sup_{[0,\frac{1}{2})}|A(d)|$. Note that

$$A(d) = \frac{2(d-d_F)+1}{L}\sum_{j=1}^{L}\left(\frac{j}{L}\right)^{2(d-d_F)}\left(\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right) \quad \text{(C.1.2)}$$

$$+\frac{2(d-d_F)+1}{L}\sum_{j=1}^{L}\left(\frac{j}{L}\right)^{2(d-d_F)}\left(\frac{\tilde{I}_{F,j}}{\tilde{g}_j}-1\right), \quad \text{(C.1.3)}$$

where $\tilde{I}_{F,j} = \gamma_L^{2d_F-1} I_{F,j}$ with $I_{F,j}$ being the periodogram of the true factor. Then (C.1.3) is equal to

$$\frac{2(d-d_F)+1}{L}\sum_{j=1}^{L}\left(\frac{j}{L}\right)^{2(d-d_F)}\left(\frac{I_{F,j}}{g_j}-1\right)$$

with $g_j = \upsilon_F \gamma_j^{-2d_F}$, whose negligibility is proved by Theorem 1 of Robinson (1995c). So we still need to show that (C.1.2) is negligible as well. By summation by parts we have

$$(C.1.2) \le \frac{3}{L}\left|\sum_{r=1}^{L-1}\left[\left(\frac{r}{L}\right)^{2(d-d_F)}-\left(\frac{r+1}{L}\right)^{2(d-d_F)}\right]\sum_{j=1}^{r}\left(\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right)\right| \quad \text{(C.1.4)}$$

$$+\frac{3}{L}\left|\sum_{j=1}^{L}\left(\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right)\right|, \quad \text{(C.1.5)}$$

where the right hand side of (C.1.4) is further bounded by

$$6\sum_{r=1}^{L-1}\left(\frac{r}{L}\right)^{1-2d_F}\frac{1}{r^2}\left|\sum_{j=1}^{r}\left(\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right)\right| \quad \text{(C.1.6)}$$

because $\left|1-\left(1+\frac{1}{r}\right)^{2(d-d_F)}\right| \le \frac{2}{r}$ on $\left[0,\frac{1}{2}\right)$ when $r>0$ and $0<1-2d_F\le 1+2(d-d_F)$ on the same range of d_F by Robinson (1995c). By Lemma C.1.2 and C.1.3, (C.1.5) is bounded by

$$\begin{aligned}
&\frac{3}{L}\sum_{j=1}^{L}\left|\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right| \\
&\le O_p(\delta_{NT})\frac{3}{L}\sum_{j=1}^{L}\left|\frac{\tilde{I}_{F,j}}{\tilde{g}_j}\right| + O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right)\frac{3}{L}\sum_{j=1}^{L}\frac{\tilde{J}_{FF,j}}{\tilde{g}_j} \\
&+O_p\left(\gamma_L^{\frac{3}{2}-3d_{F,\max}}\left\|\tilde{\delta}\right\|\right)\frac{3}{L}\sum_{j=1}^{L}\frac{\tilde{J}_{FX,j}}{\tilde{g}_j} + O_p\left(L^{-\frac{1}{2}}\gamma_L^{\frac{5}{2}-4d_{F,\max}-d_\varepsilon}\right)\frac{3}{L}\sum_{j=1}^{L}\frac{\tilde{J}_{F\varepsilon,j}}{\tilde{g}_j} \\
&+O_p\left(\gamma_L^{\frac{3}{2}-3d_{F,\max}}\right)\frac{3}{L}\sum_{j=1}^{L}\frac{\check{J}_{F\varepsilon,j}}{\tilde{g}_j} + \frac{3}{L}\sum_{j=1}^{L}\frac{I_{E_F,j}}{\tilde{g}_j} \\
&\equiv I_1+\cdots+I_5+\frac{3}{L}\sum_{j=1}^{L}\frac{I_{E_F,j}}{\tilde{g}_j},
\end{aligned} \quad \text{(C.1.7)}$$

where $\tilde{J}_{FF,j} = \left|\tilde{W}_{F,j} W^*_{F,j}\right|$, $\tilde{J}_{FX,j} = \left(\frac{1}{N}\sum_{i=1}^N \left\|\tilde{W}_{F,j} W^*_{X,ij}\right\|^2\right)^{\frac{1}{2}}$, $\tilde{J}_{F\varepsilon,j} = \left(\frac{1}{N}\sum_{i=1}^N \left|\tilde{W}_{F,j} \overline{W}_{\varepsilon,ij}\right|^2\right)^{\frac{1}{2}}$, and $\check{J}_{F\varepsilon,j} = \left\|\frac{1}{N}\sum_{i=1}^N \tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\lambda_i'\right\|$, which correspond to the scenario when $R = 1$ in Lemma C.1.3, which fits the setup in this proof. Also by Lemma C.1.3, it is sufficient to show the negligibility of I_1-I_4 because $\frac{3}{L}\sum_{j=1}^L \frac{I_{E_F,j}}{\tilde{g}_j}$ is of smaller order than the other terms. Then firstly for I_1, we have for some constant C that

$$E\left(\frac{3}{L}\sum_{j=1}^L \left|\frac{\tilde{I}_{F,j}}{\tilde{g}_j}\right|\right) = \frac{3}{L}\sum_{j=1}^L E\left|\frac{I_{F,j}}{g_j}\right| \le C \tag{C.1.8}$$

by Assumption B(iii), which is also supported by (3.16) in the proof of Theorem 1 in Robinson (1995c). Then $I_1 = o_p(1)$. Following the similar reasoning, for I_2 we have

$$E\left(\frac{3}{L}\sum_{j=1}^L \frac{\tilde{J}_{FF,j}}{\tilde{g}_j}\right) = \frac{3\gamma_L^{\frac{1}{2}-d_F}}{L}\sum_{j=1}^L E\left|\frac{I_{F,j}}{g_j}\right| = O\left(\gamma_L^{\frac{1}{2}-d_F}\right), \tag{C.1.9}$$

which then proves $I_2 = o_p(1)$. And for I_3,

$$\begin{aligned}
&E\left(\frac{3}{L}\sum_{j=1}^L \frac{\tilde{J}_{FX,j}}{\tilde{g}_j}\right) \\
&\le \frac{3}{L}\sum_{j=1}^L \frac{1}{\tilde{g}_j}\left(\frac{1}{N}\sum_{i=1}^N E\left\|\tilde{W}_{F,j} W^*_{X,ij}\right\|^2\right)^{\frac{1}{2}} \\
&\le \frac{3}{L}\sum_{j=1}^L \frac{1}{\tilde{g}_j}\left(\frac{1}{N}\sum_{i=1}^N \left(E\left|\tilde{W}_{F,j}\right|^4\right)^{\frac{1}{2}}\left(E\left\|W_{X,ij}\right\|^4\right)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\
&\le \frac{C}{L}\sum_{j=1}^L \frac{\gamma_L^{d_F-\frac{1}{2}}\gamma_j^{-d_F-d_{X,\max}}}{\tilde{g}_j} = \frac{C\gamma_L^{\frac{1}{2}-d_F}}{L}\sum_{j=1}^L \gamma_j^{d_F-d_{X,\max}} = O\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right)
\end{aligned} \tag{C.1.10}$$

by Assumption B(i), B(iii) and using Jensen's inequality, Cauchy-Schwarz inequality and Riemann-sum approximation, which altogether shows $I_3 = o_p(1)$. And next for I_4, we can follow the same reasoning as for I_3 by replacing X with ε, which implies that $E\left(\frac{3}{L}\sum_{j=1}^L \frac{\tilde{J}_{FX,j}}{\tilde{g}_j}\right) = O\left(\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$ and thus $I_4 = o_p(1)$. Lastly for I_5, we have

$$\begin{aligned}
&E\left(\frac{3}{L}\sum_{j=1}^L \frac{\check{J}_{F\varepsilon,j}}{\tilde{g}_j}\right) \\
&\le \frac{3}{L}\sum_{j=1}^L \frac{1}{\tilde{g}_j}\left(E\left\|\tilde{W}_{F,j}\right\|^2 E\left\|\frac{1}{N}\sum_{i=1}^N \overline{W}_{\varepsilon,ij}\lambda_i'\right\|^2\right)^{\frac{1}{2}} \\
&\le \frac{C}{L}\sum_{j=1}^L \frac{\gamma_L^{d_F-\frac{1}{2}}\gamma_j^{-d_F}}{\tilde{g}_j}\left(\max_{i,k}\left|E\left(\lambda_i'\lambda_k\right)\right|\frac{1}{N^2}\sum_{i,k=1}^N \left|E\left(\overline{W}_{\varepsilon,ij}\overline{W}_{\varepsilon,kj}\right)\right|\right)^{\frac{1}{2}}
\end{aligned}$$

$$\leq \frac{C}{\sqrt{N}L}\sum_{j=1}^{L}\frac{\gamma_L^{d_F-\frac{1}{2}}\gamma_j^{-d_F-d_\varepsilon}}{\tilde{g}_j} = O\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right) \tag{C.1.11}$$

by Assumption C(ii), which also proves the negligibility of I_5.

Next using the same reasoning through both sides of (C.1.7)-(C.1.11), we define

$$\Delta_{NT} = \delta_{NT} + \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} + \gamma_L^{\frac{3}{2}-3d_{F,\max}}\left\|\tilde{\delta}\right\|$$
$$+ L^{-\frac{1}{2}}\gamma_L^{\frac{5}{2}-4d_{F,\max}-d_\varepsilon} + N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-3d_{F,\max}}. \tag{C.1.12}$$

Then by changing the upper limit of summation from L to r, we can obtain a loose bound as $\sum_{j=1}^{r} E\left|\left(\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right)\right| = O\left(r\Delta_{NT}\right)$. Therefore the expectation of (C.1.6) is given by

$$6\sum_{r=1}^{L-1}\left(\frac{r}{L}\right)^{1-2d_F}\frac{1}{r^2}\sum_{j=1}^{r}E\left|\left(\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right)\right| = O\left(\Delta_{NT}\right)\frac{1}{L}\sum_{r=1}^{L-1}\left(\frac{r}{L}\right)^{-2d_F} = o(1) \tag{C.1.13}$$

by Riemann-sum approximation, which completes the proof of Theorem 4.3.1. ■

Proof of Theorem 4.3.2

Here we follow the proof of Theorem 2 in Robinson (1995c) and the notations used in the proof of Theorem 4.3.1 above. By consistency of $\hat{d}_{\hat{F}}$, as $T \to \infty$, it satisfies

$$0 = \frac{\partial K\left(\hat{d}_{\hat{F}}\right)}{\partial d} = \frac{\partial K\left(d_F\right)}{\partial d} + \frac{\partial^2 K\left(\tilde{d}\right)}{\partial d}\left(\hat{d}_{\hat{F}} - d_F\right)$$

by Taylor expansion, where $\left|\tilde{d} - d\right| \leq \left|\hat{d}_{\hat{F}} - d_F\right|$. By replacing the periodogram of true factor by the estimated one, the proof in Robinson (1995c) can proceed as

$$\frac{\partial K\left(d\right)}{\partial d} = 2\frac{\hat{G}_1(d)}{\hat{G}_0(d)} - \frac{2}{L}\sum_{j=1}^{L}\log\gamma_j, \tag{C.1.14}$$

and

$$\frac{\partial^2 K\left(d\right)}{\partial d} = \frac{4\left[\hat{G}_2(d)\hat{G}_0(d) - \hat{G}_1^2(d)\right]}{\hat{G}_0^2(d)} = \frac{4\left[\hat{F}_2(d)\hat{F}_0(d) - \hat{F}_1^2(d)\right]}{\hat{F}_0^2(d)}$$
$$= \frac{4\left[\hat{E}_2(d)\hat{E}_0(d) - \hat{E}_1^2(d)\right]}{\hat{E}_0^2(d)},$$

where

$$\hat{G}_k(d) = \frac{\gamma_L^{1-2d_F}}{L}\sum_{j=1}^{L}\left(\log\gamma_j\right)^k\gamma_j^{2d}\hat{I}_{F,j},$$

and

$$\hat{F}_k(d) = \frac{\gamma_L^{1-2d_F}}{L}\sum_{j=1}^{L}(\log j)^k\,\gamma_j^{2d}\hat{I}_{F,j},\text{ and }\hat{E}_k(d) = \frac{\gamma_L^{1-2d_F}}{L}\sum_{j=1}^{L}(\log j)^k\,j^{2d}\hat{I}_{F,j}.$$

As in Robinson (1995c), we fix $\varepsilon > 0$ such that $2\varepsilon < (\log L)^2$, then we define the set $M = \left\{d : (\log L)^3\,|d - d_F| \le \varepsilon\right\}$, in which

$$\left|\hat{E}_k(d) - \hat{E}_k(d_F)\right| \le \frac{\gamma_L^{1-2d_F}}{L}\sum_{j=1}^{L}\left|j^{2(d-d_F)} - 1\right|(\log j)^k\,j^{2d_F}\hat{I}_{F,j} \le 2e\varepsilon\,(\log L)^{k-2}\,\hat{E}_0(d_F)\,. \tag{C.1.15}$$

The above inequality holds because the reasoning required are all about the deterministic part of $\hat{E}_k(d)$. Therefore for $\eta > 0$,

$$\begin{aligned}&P\left(\left|\hat{E}_k\left(\tilde{d}\right) - \hat{E}_k(d_F)\right| > \eta\left(\frac{2\pi}{T}\right)^{-2d_F}\right)\\ &\le P\left(\hat{G}_0(d_F) > \frac{\eta}{2e\varepsilon}(\log L)^{2-k}\right) + P\left((\log L)^3\,|d - d_F| > \varepsilon\right).\end{aligned} \tag{C.1.16}$$

For $k = 0, 1, 2$, the first probability in (C.1.16) tends to zero because $\hat{E}_0(d_F) = \left(\frac{2\pi}{T}\right)^{-2d_F}\hat{G}_0(d_F)$ and $\hat{G}_0(d_F) \xrightarrow{p} \upsilon_F \in (0,\infty)$ according to the asymptotic negligibility of $\sup_{[0,\frac{1}{2})}|A(d)|$ as given in the proof of Theorem 4.3.1. Continue following Robinson (1995c) as we did before, to prove the negligibility of second probability, it is sufficient to show

$$P\left(\inf_{\bar{N}_\delta\cap[0,\frac{1}{2})\cap\overline{M}} U(d) \le \sup_{\bar{N}_\delta\cap[0,\frac{1}{2})}|V(d)|\right) \to 0, \tag{C.1.17}$$

where $\overline{M} = (-\infty,\infty) - M$. Based on the lower bound of $U(d)$ under the inclusion of $\overline{M}$, and the reasoning about the order of $V(d)$ in the proof of Theorem 4.3.1, (C.1.17) holds if $\sup_{\bar{N}_\delta\cap[0,\frac{1}{2})}|A(d)| = o_p\left((\log L)^{-6}\right)$. Recall from (C.1.2) and (C.1.3) that

$$\begin{aligned}A(d) &= \frac{2(d-d_F)+1}{L}\sum_{j=1}^{L}\left(\frac{j}{L}\right)^{2(d-d_F)}\left(\frac{\hat{I}_{F,j} - \tilde{I}_{F,j}}{\tilde{g}_j}\right)\\ &+ \frac{2(d-d_F)+1}{L}\sum_{j=1}^{L}\left(\frac{j}{L}\right)^{2(d-d_F)}\left(\frac{\tilde{I}_{F,j}}{\tilde{g}_j} - 1\right) \equiv A_1(d) + A_2(d),\end{aligned}$$

we can see Robinson (1995c) has already proved $\sup_{\bar{N}_\delta\cap[0,\frac{1}{2})}|A_2(d)| = o_p\left((\log L)^{-6}\right)$, and the orders given by the left hand side of (C.1.7) shows $\sup_{\bar{N}_\delta\cap[0,\frac{1}{2})}|A_1(d)| \le \sup_{[0,\frac{1}{2})}|A_1(d)| = o_p\left((\log L)^{-6}\right)$, which altogether proves (C.1.17) and the negligibility of the left hand side of (C.1.16). Such conclusion, as shown in Robinson (1995c), implies

$$\frac{\partial^2 K\left(\tilde{d}\right)}{\partial d} = \frac{4\left[\hat{F}_2\left(d_F\right)\hat{F}\left(d_F\right) - \hat{F}_1^2\left(d_F\right)\right]}{\hat{F}_0^2\left(d_F\right)} + o_p(1)$$

as $T \to \infty$. Next for $k \geq 0$,

$$\begin{aligned}
&\left|\hat{F}_k\left(d_F\right) - \upsilon_F \frac{1}{L}\sum_{j=1}^{L}\left(\log j\right)^k\right| \\
&= \left|\frac{1}{L}\sum_{j=1}^{L}\left(\log j\right)^k\left[\gamma_L^{1-2d_F}\gamma_j^{2d_F}\hat{I}_{F,j} - \upsilon_F\right]\right| = \left|\frac{\upsilon_F}{L}\sum_{j=1}^{L}\left(\log j\right)^k\left[\frac{\hat{I}_{F,j}}{\tilde{g}_j} - 1\right]\right| \\
&\leq \frac{\upsilon_F}{L}\sum_{j=1}^{L}\left(\log j\right)^k\left|\frac{\hat{I}_{F,j} - \tilde{I}_{F,j}}{\tilde{g}_j}\right| + \frac{\upsilon_F}{L}\sum_{j=1}^{L}\left(\log j\right)^k\left|\frac{I_{F,j}}{g_j} - 1\right| \equiv F_1 + F_2,
\end{aligned}$$

where Robinson (1995c) has proved the asymptotic negligibility of F_2. And for F_1, we have by summation by parts that

$$\begin{aligned}
&\frac{\upsilon_F}{L}\sum_{j=1}^{L}\left(\log j\right)^k\left|\frac{\hat{I}_{F,j} - \tilde{I}_{F,j}}{\tilde{g}_j}\right| \\
&\leq \frac{\upsilon_F}{L}\sum_{r=1}^{L-1}\left|\left(\log r\right)^k - \left(\log(r+1)\right)^k\right|\left|\sum_{j=1}^{r}\frac{\hat{I}_{F,j} - \tilde{I}_{F,j}}{\tilde{g}_j}\right| + \frac{\upsilon_F}{L}\left(\log L\right)^k\left|\sum_{j=1}^{L}\frac{\hat{I}_{F,j} - \tilde{I}_{F,j}}{\tilde{g}_j}\right| \\
&\equiv F_{11} + F_{12}.
\end{aligned}$$

Firstly for F_{12}, the reasoning behind (C.1.7) based on Lemma C.1.3 continues to hold and proves its negligibility. Next for F_{11}, we adopt the conclusion in Robinson (1995c) that $\left|\left(\log r\right)^k - \left(\log(r+1)\right)^k\right| \leq \frac{\left(\log(r+1)\right)^{k-1}}{r}$. Note that following the reasoning as we analyze (C.1.6), we have

$$\frac{1}{L}\sum_{r=1}^{L-1}\left(\log(r+1)\right)^{k-1}\frac{1}{r}\sum_{j=1}^{r}E\left|\left(\frac{\hat{I}_{F,j} - \tilde{I}_{F,j}}{\tilde{g}_j}\right)\right| = O\left(\left(\Delta_{NT}\right)\left(\log L\right)^{k-1}\right),$$

which corresponds to (C.1.13) and is negligible because it is dominated by F_{12}. Therefore, identical to what Robinson (1995c) has obtained, the limit of F_2 dominates in the limit and thus $\frac{\partial^2 K\left(\tilde{d}\right)}{\partial d} \xrightarrow{p} 4$.

Next we turn to $\frac{\partial K(d_F)}{\partial d}$. By (C.1.14) and the fact that $\hat{G}_0\left(d_F\right) \xrightarrow{p} \upsilon_F$,

$$\begin{aligned}
L^{\frac{1}{2}}\frac{\partial K\left(d_F\right)}{\partial d} &= 2L^{-\frac{1}{2}}\sum_{j=1}^{L}\left(\log\gamma_j\right)\frac{\gamma_L^{1-2d_F}\gamma_j^{2d_F}\hat{I}_{F,j}}{\hat{G}_0\left(d_F\right)} - 2L^{-\frac{1}{2}}\sum_{j=1}^{L}\log\gamma_j \\
&= 2L^{-\frac{1}{2}}\sum_{j=1}^{L}u_j\frac{\gamma_L^{1-2d_F}\gamma_j^{2d_F}\hat{I}_{F,j}}{\upsilon_F + o_p(1)}
\end{aligned}$$

$$
\begin{aligned}
&= 2L^{-\frac{1}{2}}\left(1-\frac{o_p(1)}{\upsilon_F+o_p(1)}\right)\sum_{j=1}^{L}u_j\frac{\hat{I}_{F,j}}{\tilde{g}_j}\\
&= 2L^{-\frac{1}{2}}\sum_{j=1}^{L}u_j\left(\frac{\hat{I}_{F,j}}{\tilde{g}_j}-1\right)(1+o_p(1))\\
&= 2L^{-\frac{1}{2}}\sum_{j=1}^{L}u_j\left(\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right)(1+o_p(1))+2L^{-\frac{1}{2}}\sum_{j=1}^{L}u_j\left(\frac{I_{F,j}}{g_j}-1\right)(1+o_p(1))\\
&\equiv L_1+L_2, \qquad \text{(C.1.18)}
\end{aligned}
$$

where $u_j = \log j - \frac{1}{L}\sum_{j=1}^{L}\log j$ and the second last equality holds by the fact that $\sum_{j=1}^{L}u_j = 0$. Firstly for L_1, we have

$$
L^{-\frac{1}{2}}\sum_{j=1}^{L}u_jE\left|\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right| \leq \left(\sqrt{L}\log L\right)\frac{1}{L}\sum_{j=1}^{L}E\left|\frac{\hat{I}_{F,j}-\tilde{I}_{F,j}}{\tilde{g}_j}\right| = O\left(\Delta_{NT}\sqrt{L}\log L\right),
$$

which is negligible by the definition of Δ_{NT} in (C.1.12) by Assumption E(i). And lastly we can see that L_2 is identical to (4.11) in the proof of Theorem 2 in Robinson (1995c), which follows that $L_2 \xrightarrow{d} \mathcal{N}(0,4)$, and thus the proof of Theorem 4.3.2 is completed. ■

Proof of Theorem 4.4.1

Proof of this theorem is based on the proof of Theorem 1 in Qu (2011). And same as before, such statistic is calculated marginally at each $r = 1, \ldots, R$ in the same manner. So we omit the subscript r as we did in the proof of the theorems above and treat the factor as a scalar. Since by construction $\frac{1}{L}\sum_{j=1}^{L}u_j^2 \to 1$, it is sufficient to consider the uniform convergence of

$$
\tilde{V}_L\left(\rho,\hat{d}_{\hat{F}}\right) = L^{-\frac{1}{2}}\sum_{j=1}^{\lfloor L\rho\rfloor}u_j\left(\frac{\hat{I}_{F,j}}{\hat{G}\left(\hat{d}_{\hat{F}}\right)\gamma_j^{-2\hat{d}_{\hat{F}}}}-1\right)
$$

over $\rho \in [0,1]$. To proceed, we conduct the first-order Taylor expansion at $\hat{d}_{\hat{F}} = d_F$,

$$
\tilde{V}_L\left(\rho,\hat{d}_{\hat{F}}\right) = L^{-\frac{1}{2}}\sum_{j=1}^{\lfloor L\rho\rfloor}u_j\left(\frac{\hat{I}_{F,j}}{\hat{G}\left(d_F\right)\gamma_j^{-2d_F}}-1\right)+L^{-\frac{1}{2}}\frac{\partial\tilde{V}_L\left(\rho,\tilde{d}\right)}{\partial d}L^{\frac{1}{2}}\left(\hat{d}_{\hat{F}}-d_F\right)\equiv V_1+V_2,
$$

where

$$
L^{-\frac{1}{2}}\frac{\partial\tilde{V}_L\left(\rho,\tilde{d}\right)}{\partial d} = \frac{2}{L\hat{G}\left(\tilde{d}\right)^2}\sum_{j=1}^{\lfloor L\rho\rfloor}u_j\frac{\hat{I}_{F,j}}{\gamma_j^{-2\tilde{d}}}\left\{\frac{1}{L}\log\gamma_j\sum_{k=1}^{L}\frac{\hat{I}_{F,k}}{\gamma_k^{-2\tilde{d}}}-\frac{1}{L}\sum_{s=1}^{L}\log\gamma_s\frac{\hat{I}_{F,s}}{\gamma_s^{-2\tilde{d}}}\right\}.
$$

Following the reasoning on both sides of (C.1.18), we have

$$V_1 = L^{-\frac{1}{2}} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j \left(\frac{I_{F,j}}{g_j} - 1 \right) (1 + o_p(1)) + o_p(1)$$

with $g_j = \upsilon_F \gamma_j^{-2d_F}$. Such conclusion holds because by arguments like (C.1.13) the above approximation can be applied at any upper limit of summation less than or equal to L, therefore also applied uniformly in $\rho \in [0,1]$. This implies (A.2) in the proof of Theorem 1 in Qu (2011) equals to V_1 asymptotically.

Next for V_2, by replacing $I_{F,j}$ with its FDPCLS estimator, and using the notation of $\tilde{\upsilon}_F$ and $\tilde{G}(d)$ in the proof of Theorem 4.3.1 and 4.3.2, the reasoning in Qu (2011) continues to hold as we can rewrite it as

$$V_2 = \frac{2}{\hat{G}\left(\tilde{d}\right)^2} \left(L^{-1} \sum_{k=1}^{L} \frac{\hat{I}_{F,k}}{\gamma_k^{-2\tilde{d}}} \right) \left(L^{-1} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^2 \frac{\hat{I}_{F,j}}{\gamma_j^{-2\tilde{d}}} \right) - \frac{2}{\hat{G}\left(\tilde{d}\right)^2} \left(L^{-1} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j \frac{\hat{I}_{F,j}}{\gamma_j^{-2\tilde{d}}} \right) \left(L^{-1} \sum_{s=1}^{L} u_s \frac{\hat{I}_{F,s}}{\gamma_s^{-2\tilde{d}}} \right).$$

In the following we try to analyze the asymptotics of $\frac{1}{\hat{G}(\tilde{d})} \left(L^{-1} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^k \frac{\hat{I}_{F,j}}{\gamma_j^{-2\tilde{d}}} \right)$ with $k = 0, 1, 2$, using the similar approximation scheme as we used before. To be specific,

$$\frac{1}{\hat{G}\left(\tilde{d}\right)} \left(L^{-1} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^k \frac{\hat{I}_{F,j}}{\gamma_j^{-2\tilde{d}}} \right) = \frac{L^{-1} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^k \gamma_j^{2\tilde{d}} \hat{I}_{F,j}}{L^{-1} \sum_{j=1}^{L} \gamma_j^{2\tilde{d}} \hat{I}_{F,j}} = \frac{\frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^k j^{2\tilde{d}} \hat{I}_{F,j}}{\frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{L} j^{2\tilde{d}} \hat{I}_{F,j}}, \quad \text{(C.1.19)}$$

where the numerator of the right hand side of (C.1.19), under the case $k = 0, 1, 2$, are further given respectively by

$$\frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} j^{2\tilde{d}} \hat{I}_{F,j} \equiv \hat{E}_0\left(\rho, \tilde{d}\right), \text{ when } k = 0;$$

$$\begin{aligned} \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j j^{2\tilde{d}} \hat{I}_{F,j} &= \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} (\log j)\, j^{2\tilde{d}} \hat{I}_{F,j} - \left(\frac{1}{L} \sum_{j=1}^{L} \log j \right) \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} j^{2\tilde{d}} \hat{I}_{F,j} \\ &\equiv \hat{E}_1\left(\rho, \tilde{d}\right) - \left(\frac{1}{L} \sum_{j=1}^{L} \log j \right) \hat{E}_0\left(\rho, \tilde{d}\right), \text{ when } k = 1; \end{aligned}$$

and

$$\frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^2 j^{2\tilde{d}} \hat{I}_{F,j} = \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} (\log j)^2 j^{2\tilde{d}} \hat{I}_{F,j} - \left(\frac{2}{L} \sum_{j=1}^{L} \log j \right) \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} (\log j)\, j^{2\tilde{d}} \hat{I}_{F,j}$$

$$+ \left(\frac{1}{L} \sum_{j=1}^{L} \log j \right)^2 \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} j^{2\tilde{d}} \hat{I}_{F,j}$$
$$\equiv \hat{E}_2 \left(\rho, \tilde{d}\right) - \left(\frac{2}{L} \sum_{j=1}^{L} \log j \right) \hat{E}_1 \left(\rho, \tilde{d}\right) + \left(\frac{1}{L} \sum_{j=1}^{L} \log j \right)^2 \hat{E}_0 \left(\rho, \tilde{d}\right), \text{ when } k = 2.$$

Therefore it is sufficient to analyze the asymptotics of $\hat{E}_k \left(\rho, \tilde{d}\right)$ for $k = 0, 1, 2$. To proceed, we denote $\delta \left(\hat{d}_{\hat{F}}, d_F\right)$ as the interval between $\hat{d}_{\hat{F}}$ and d_F, and firstly try to prove

$$\sup_{d \in \delta\left(\hat{d}_{\hat{F}}, d_F\right)} \sup_{\rho \in [0,1]} \left| \hat{E}_k (\rho, d) - \hat{E}_k (\rho, d_F) \right| = o_p \left(T^{2d_F} (\log L)^{-2} \right).$$

To be specific, we can follow the reasoning in (C.1.15) and (C.1.16) because

$$\sup_{d \in \delta\left(\hat{d}_{\hat{F}}, d_F\right)} \sup_{\rho \in [0,1]} \left| \hat{E}_k (\rho, d) - \hat{E}_k (\rho, d_F) \right|$$
$$= \sup_{d \in \delta\left(\hat{d}_{\hat{F}}, d_F\right)} \sup_{\rho \in [0,1]} \left| \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} (\log j)^k \left(j^{2d} - j^{2d_F} \right) \hat{I}_{F,j} \right|$$
$$\leq \sup_{d \in \delta\left(\hat{d}_{\hat{F}}, d_F\right)} \frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{L} \left| j^{2(d-d_F)} - 1 \right| (\log j)^k j^{2d_F} \hat{I}_{F,j},$$

which can be bounded as in (C.1.15) and (C.1.16) by defining the alternative set given by $M = \left\{ d : (\log L)^5 |d - d_F| \leq \varepsilon \right\}$ as given by the proof of Lemma B.2 in Qu (2011), which is also indicated by Andrews and Sun (2004) and Wu and Shao (2007), and thus can be proved as $o_p \left(T^{2d_F} \right)$ as indicated by the left hand side of (C.1.16). Next we try to prove

$$\sup_{\rho \in [0,1]} \left| \hat{E}_k (\rho, d_F) - E_k (\rho, d_F) \right| = o_p \left(T^{2d_F} (\log L)^{-2} \right),$$

where $E_k (\rho, d_F) = \frac{1}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} (\log j)^k j^{2d_F} I_{F,j}$. To be specific,

$$\sup_{\rho \in [0,1]} \left| \hat{E}_k (\rho, d_F) - E_k (\rho, d_F) \right| = \sup_{\rho \in [0,1]} \left| \frac{1}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} (\log j)^k j^{2d_F} \left(\gamma_L^{1-2d_F} \hat{I}_{F,j} - I_{F,j} \right) \right|$$
$$\leq \left(\frac{2\pi}{T} \right)^{-2d_F} \frac{1}{L} \sum_{j=1}^{L} (\log j)^k \left| \frac{\hat{I}_{F,j} - \tilde{I}_{F,j}}{\tilde{g}_j} \right| = o_p \left(T^{2d_F} (\log L)^{-2} \right)$$

by how we prove the negligibility of F_1 in the proof of Theorem 4.3.2. Now combining what we have obtained so far, the right hand side of (C.1.19) follows that

$$\frac{\frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^k j^{2\tilde{d}} \hat{I}_{F,j}}{\frac{\gamma_L^{1-2d_F}}{L} \sum_{j=1}^{L} j^{2\tilde{d}} \hat{I}_{F,j}} = \frac{\frac{1}{L} \sum_{j=1}^{\lfloor L\rho \rfloor} u_j^k j^{2d_F} I_{F,j} + o_p \left(T^{2d_F} \right)}{\frac{1}{L} \sum_{j=1}^{L} j^{2d_F} I_{F,j} + o_p \left(T^{2d_F} \right)},$$

then the proof of Lemma B.3 in Qu (2011) will hold and implies that

$$\frac{1}{\hat{G}\left(\tilde{d}\right)}\left(L^{-1}\sum_{j=1}^{\lfloor L\rho\rfloor}u_j^k\frac{\hat{I}_{F,j}}{\gamma_j^{-2\tilde{d}}}\right)=\frac{1}{L}\sum_{j=1}^{\lfloor L\rho\rfloor}u_j^k+o_p(1),$$

then the proof of Theorem 1 in Qu (2011) continues to hold as $V_2\to 2\Phi(\rho)$. Combining all we have so far, the expression (A.4) in Qu (2011) holds asymptotically for our $\tilde{V}_L\left(\rho,\hat{d}_{\hat{F}}\right)$ as

$$\tilde{V}_L\left(\rho,\hat{d}_{\hat{F}}\right)=\left[L^{-\frac{1}{2}}\frac{\upsilon_F}{G(d_F)}\sum_{j=1}^{\lfloor L\rho\rfloor}u_j\left(\frac{I_{F,j}}{\upsilon_F\gamma_j^{-2d_F}}-1\right)-L^{-\frac{3}{2}}\frac{\upsilon_F}{G(d_F)}\left(\sum_{j=1}^{\lfloor L\rho\rfloor}u_j\right)\sum_{j=1}^{L}\left(\frac{I_{F,j}}{\upsilon_F\gamma_j^{-2d_F}}-1\right)\right]$$
$$\times(1+o_p(1))+2\Phi(\rho)L^{\frac{1}{2}}\left(\hat{d}_{\hat{F}_r}-d_F\right)+o_p(1).$$

Therefore the finite-sample convergence and the tightness will hold as in Qu (2011), which completes the proof of our Theorem 4.4.1. ■

C.2 Proofs of the Technical Lemmas

Proof of Lemma C.1.2

Proof of this lemma will closely follow the proof of (2) in Bai and Ng (2013), as it is adopted using our framework of linear regression with interactive fixed effects. To be specific,

$$\frac{\tilde{W}_F^*\hat{W}_F}{T}=\frac{1}{T}\tilde{W}_F^*\left(\hat{W}_F-\tilde{W}_F\tilde{H}\right)+\frac{1}{T}\tilde{W}_F^*\tilde{W}_F\tilde{H}. \tag{C.2.1}$$

Note that by Lemma C.1.1 and Assumption B(iii), $\frac{1}{T}\tilde{W}_F^*\left(\hat{W}_F-\tilde{W}_F\tilde{H}\right)=O_p\left(\delta_{NT}\right)$ in which

$$\delta_{NT}=\delta_{W1,NT}\left(\tilde{\beta}-\beta\right)+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon},$$

where $\delta_{W1,NT}=\gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}}+\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$ as presented in Lemma C.1.1. Then left multiplying $\tilde{H}^*$ to both sides of (C.2.1), we have

$$\frac{1}{T}\tilde{H}^*\tilde{W}_F^*\hat{W}_F=\frac{1}{T}\tilde{H}^*\tilde{W}_F^*\tilde{W}_F\tilde{H}+O_p\left(\delta_{(1),NT}\right), \tag{C.2.2}$$

where $\delta_{(1),NT}=\gamma_L^{1-2d_{F,\max}}\delta_{NT}$ because $\tilde{H}=O_p\left(\gamma_L^{1-2d_{F,\max}}\right)$ by construction. Furthermore the left hand side of (C.2.2) is given by

$$\frac{1}{T}\tilde{H}^*\tilde{W}_F^*\hat{W}_F=\frac{1}{T}\left(\tilde{H}^*\tilde{W}_F^*-\hat{W}_F^*+\hat{W}_F^*\right)\hat{W}_F$$

$$= I + O_p(\delta_{NT}) \tag{C.2.3}$$

by the identifying restriction of our FDPCLS estimator and Lemma C.1.1. Equating (C.2.3) and the right hand side of (C.2.2) we have

$$\frac{1}{T}\tilde{H}^* \tilde{W}_F^* \tilde{W}_F \tilde{H} = \tilde{H}^* \tilde{H} = I + O_p(\delta_{NT})$$

by Assumption F(i). Following the same reasoning as in Bai and Ng (2013), we can ignore the negligible term $O_p(\delta_{NT})$, and $\tilde{H}$ is a unitary matrix, the complex generalization of orthogonal matrix, which means $\tilde{H}$ has its eigenvalues being either 1 or -1. Then in the following it is sufficient to prove that $\tilde{H}$ is also a diagonal matrix, after which we can assume all its eigenvalues are 1 by multiplying the columns by -1 of $\hat{W}_F$ corresponding to -1 eigenvalues.

Recall that $\tilde{H}$ by definition is given by

$$\begin{aligned}
\check{H} &= \left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\left(\frac{\tilde{W}_F^* \hat{W}_F}{T}\right) V_{NL}^{-1} = \left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\left(\frac{1}{T}\tilde{W}_F^* \tilde{W}_F \tilde{H} + \frac{1}{T}\tilde{W}_F^*\left(\hat{W}_F - \tilde{W}_F \tilde{H}\right)\right) V_{NL}^{-1} \\
&= \left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\tilde{H} V_{NL}^{-1} + O_p\left(\delta_{(1),NT}\right) \equiv A\tilde{H} V_{NL}^{-1} + O_p\left(\delta_{(1),NT}\right) \qquad \text{(C.2.4)}
\end{aligned}$$

by Assumption F(i) and the construction of $\tilde{\Lambda}$. Then right multiplying V_{NL} on both sides of (C.2.4) we have

$$A\tilde{H} = \tilde{H} V_{NL} + O_p\left(\delta_{(1),NT}\right).$$

This replicates the form of (23) in Bai and Ng (2013) except that $\tilde{H}$ could be a complex matrix. But it can still be a matrix of eigenvectors of matrix A as the columns of $\tilde{H}$ now form the eigenbasis of a complex space. Then the rest of reasoning in Bai and Ng (2013) will hold because by Assumption F(ii), which implies that A is a diagonal matrix with distinct eigenvalues. Therefore $\tilde{\tilde{H}} = I + O_p(\delta_{NT})$. ■

Proof of Lemma C.1.3

From the right hand side of both (4.3.1) and (4.3.2), we have for each $j = 1, \ldots, L$ that

$$\hat{W}_{F,j} = \left[\frac{1}{NT}\sum_{i=1}^{N} \hat{W}_{U,ij} \hat{W}_{U,i}^*\right] \hat{W}_F V_{NL}^{-1}, \tag{C.2.5}$$

where we denote $\hat{W}_{U,i} = W_{Y,i} - W_{X,i}\tilde{\beta}$. Also we denote $\tilde{\delta} = \tilde{\beta} - \beta$. By expanding $\hat{W}_{U,i}$ as denoted just now and $W_{Y,i}$ by (4.3.3), (C.2.5) becomes

$$\begin{aligned}
\hat{W}'_{F,j} =& \frac{1}{NT}\sum_{i=1}^{N} W'_{X,ij}\tilde{\delta}\tilde{\delta}' W^*_{X,i}\hat{W}_F V_{NL}^{-1} - \frac{1}{NT}\sum_{i=1}^{N} W'_{X,ij}\tilde{\delta}\lambda'_i W^*_F \hat{W}_F V_{NL}^{-1} - \frac{1}{NT}\sum_{i=1}^{N} W'_{X,ij}\tilde{\delta} W^*_{\varepsilon,i}\hat{W}_F V_{NL}^{-1} \\
& - \frac{1}{NT}\sum_{i=1}^{N} W'_{F,j}\lambda_i\tilde{\delta}' W^*_{X,i}\hat{W}_F V_{NL}^{-1} - \frac{1}{NT}\sum_{i=1}^{N} W_{\varepsilon,ij}\tilde{\delta}' W^*_{X,i}\hat{W}_F V_{NL}^{-1} + \frac{1}{NT}\sum_{i=1}^{N} W'_{F,j}\lambda_i W^*_{\varepsilon,i}\hat{W}_F V_{NL}^{-1} \\
& + \frac{1}{NT}\sum_{i=1}^{N} W_{\varepsilon,ij}\lambda'_i W^*_F \hat{W}_F V_{NL}^{-1} + \frac{1}{NT}\sum_{i=1}^{N} W_{\varepsilon,ij} W^*_{\varepsilon,i}\hat{W}_F V_{NL}^{-1} + \frac{1}{NT}\sum_{i=1}^{N} \tilde{W}'_{F,j}\tilde{\lambda}_i\tilde{\lambda}'_i \tilde{W}^*_F \hat{W}_F V_{NL}^{-1} \\
& \equiv \tilde{I}_1 + \cdots + \tilde{I}_9.
\end{aligned}$$

Since $\tilde{I}_9 = \tilde{W}'_{F,j}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)\left(\tilde{W}_F\hat{W}^*_F/T\right)V_{NL}^{-1}$, we have

$$\hat{W}'_{F,j} - \tilde{W}'_{F,j}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)\left(\tilde{W}_F\hat{W}^*_F/T\right)V_{NL}^{-1} = \tilde{I}_1 + \cdots + \tilde{I}_8 = \hat{W}'_{F,j} - \tilde{W}'_{F,j}\tilde{H} \equiv E'_{F,j}. \quad \text{(C.2.6)}$$

From (C.2.6) we have

$$\hat{W}'_{F,j} = \tilde{W}'_{F,j} + \tilde{W}'_{F,j}\left(\tilde{H} - I\right) + E'_{F,j},$$

then the estimated periodogram $\hat{I}_{F,j} = \hat{W}_{F,j}\hat{W}^*_{F,j}$ follows that

$$\begin{aligned}
\hat{I}_{F,j} =& \tilde{I}_{F,j} + \tilde{I}_{F,j}\overline{\left(\tilde{H} - I\right)} + \tilde{W}_{F,j}E^*_{F,j} + \left(\tilde{H} - I\right)'\tilde{I}_{F,j} + \left(\tilde{H} - I\right)'\tilde{I}_{F,j}\overline{\left(\tilde{H} - I\right)} \\
& + \left(\tilde{H} - I\right)'\tilde{W}_{F,j}E^*_{F,j} + E_{F,j}\tilde{W}^*_{F,j} + E_{F,j}W^*_{F,j}\overline{\left(\tilde{H} - I\right)} + I_{E_F,j} \\
& = \tilde{I}_{F,j} + \tilde{I}_{F,j}\overline{\left(\tilde{H} - I\right)} + \left(\tilde{H} - I\right)'\tilde{I}_{F,j} + \left(\tilde{H} - I\right)'\tilde{I}_{F,j}\overline{\left(\tilde{H} - I\right)} + \tilde{H}'\tilde{W}_{F,j}E^*_{F,j} + E_{F,j}\tilde{W}^*_{F,j}\overline{\tilde{H}} + I_{E_F,j} \\
& \equiv \tilde{I}_{F,j} + \tilde{J}_1 + \cdots + \tilde{J}_6,
\end{aligned}$$

where $\tilde{I}_{F,j} = \tilde{W}_{F,j}\tilde{W}^*_{F,j}$, and $I_{E_F,j}$ is defined in the same manner. Then in the following we derive the orders of $\tilde{J}_1, \ldots, \tilde{J}_6$, taking the structure of $E_{F,j}$ given by (C.2.6) into consideration and following the proof of Lemma A.6 in our chapter 2. Firstly for $\tilde{J}_1$, we have

$$\begin{aligned}
\left\|\tilde{J}_1\right\| = \left\|\tilde{I}_{F,j}\overline{\left(\tilde{H} - I\right)}\right\| &\leq O_p\left(\delta_{NT}\right)\left\|\tilde{I}_{F,j}\right\| \\
&= O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\left\|\tilde{I}_{F,j}\right\|
\end{aligned}$$

by Lemma C.1.2, and the same order holds for $\tilde{J}_2$. Similarly $\tilde{J}_3 \leq O_p\left(\delta^2_{NT}\right)\left\|\tilde{I}_{F,j}\right\|$ and thus it will be dominated by $\tilde{J}_1$ and $\tilde{J}_2$. Therefore we set them as parts of the leading terms later. Next for $\tilde{J}_4$, by substituting (C.2.6) we have

$$\tilde{J}_4 = \tilde{H}'\tilde{W}_{F,j}E^*_{F,j} = \tilde{H}'\tilde{W}_{F,j}\overline{\left(\tilde{I}_1 + \cdots + \tilde{I}_8\right)}$$

$$
\begin{aligned}
&= \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{X,ij}\tilde{\delta}\tilde{\delta}'W'_{X,i}\overline{\hat{W}}_F V_{NL}^{-1} - \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{X,ij}\tilde{\delta}\lambda_i'W_F'\overline{\hat{W}}_F V_{NL}^{-1} \\
&- \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{X,ij}\tilde{\delta}W'_{\varepsilon,i}\overline{\hat{W}}_F V_{NL}^{-1} - \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{F,j}\lambda_i\tilde{\delta}'W'_{X,i}\overline{\hat{W}}_F V_{NL}^{-1} \\
&- \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\tilde{\delta}'W'_{X,i}\overline{\hat{W}}_F V_{NL}^{-1} + \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{F,j}\lambda_i W'_{\varepsilon,i}\overline{\hat{W}}_F V_{NL}^{-1} \\
&+ \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\lambda_i'W_F'\overline{\hat{W}}_F V_{NL}^{-1} + \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}W'_{\varepsilon,i}\overline{\hat{W}}_F V_{NL}^{-1} \\
&\equiv \tilde{J}_{4,1} + \cdots + \tilde{J}_{4,8}.
\end{aligned}
$$

We first analyze the order of $\left\|\tilde{J}_{4,1}\right\|$, which follows that

$$
\begin{aligned}
\left\|\tilde{J}_{4,1}\right\| &= \left\|\tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{X,ij}\tilde{\delta}\tilde{\delta}'W'_{X,i}\overline{\hat{W}}_F V_{NL}^{-1}\right\| \\
&\leq \left\|\tilde{H}\right\|\left\|\tilde{\delta}\right\|^2\frac{1}{NT}\sum_{i=1}^{N}\left\|\tilde{W}_{F,j}W^*_{X,ij}\right\|\left\|W_{X,i}\right\|\left\|\hat{W}_F\right\|\left\|V_{NL}^{-1}\right\| \\
&\leq O_p\left(\gamma_L^{1-2d_{F,\max}}\left\|\tilde{\delta}\right\|^2\right)\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\left\|\tilde{W}_{F,j}W^*_{X,ij}\right\|\left\|W_{X,i}\right\| \\
&\leq O_p\left(\gamma_L^{1-2d_{F,\max}}\left\|\tilde{\delta}\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{W}_{F,j}W^*_{X,ij}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{i=1}^{N}\left\|W_{X,i}\right\|^2\right)^{\frac{1}{2}} \\
&\leq O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{W}_{F,j}W^*_{X,ij}\right\|^2\right)^{\frac{1}{2}} \\
&\equiv O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\|^2\right)\tilde{J}_{FX,j}
\end{aligned}
$$

by Cauchy-Schwarz inequality and Assumption B(i). Then by the same reasoning as we analyze $\tilde{J}_{4,1}$ where we follow the reasoning in the proof of Lemma A.6 in our chapter two, we can conclude that

$$
\left\|\tilde{J}_{4,2}\right\| = \left\|\tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{X,ij}\tilde{\delta}\lambda_i'W_F'\overline{\hat{W}}_F V_{NL}^{-1}\right\| \leq O_p\left(\gamma_L^{\frac{3}{2}-3d_{F,\max}}\left\|\tilde{\delta}\right\|\right)\tilde{J}_{FX,j},
$$

$$
\left\|\tilde{J}_{4,3}\right\| = \left\|\tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{X,ij}\tilde{\delta}W'_{\varepsilon,i}\overline{\hat{W}}_F V_{NL}^{-1}\right\| \leq O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|\right)\tilde{J}_{FX,j},
$$

$$
\begin{aligned}
\left\|\tilde{J}_{4,4}\right\| &= \left\|\tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{F,j}\lambda_i\tilde{\delta}'W'_{X,i}\overline{\hat{W}}_F V_{NL}^{-1}\right\| \leq O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\|\right)\left\|\tilde{W}_{F,j}W^*_{F,j}\right\| \\
&\equiv O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\|\right)\tilde{J}_{FF,j}
\end{aligned}
$$

and

$$\left\|\tilde{J}_{4,5}\right\| = \left\|\tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\tilde{\delta}'W'_{X,i}\overline{\hat{W}}_F V_{NL}^{-1}\right\| \leq O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\|\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\right\|^2\right)^{\frac{1}{2}}$$
$$\equiv O_p\left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}}\left\|\tilde{\delta}\right\|\right)\tilde{J}_{F\varepsilon,j}.$$

Next for $\tilde{J}_{4,6}$, we have

$$\left\|\tilde{J}_{4,6}\right\| = \left\|\tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}W^*_{F,j}\lambda_i W'_{\varepsilon,i}\overline{\hat{W}}_F V_{NL}^{-1}\right\| \lesssim \left\|\tilde{H}\right\|\left\|\tilde{W}_{F,j}W^*_{F,j}\right\|\frac{1}{N\sqrt{T}}\left\|\sum_{i=1}^{N}\lambda_i W'_{\varepsilon,i}\right\|$$
$$= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{\varepsilon}}\right)\tilde{J}_{FF,j}$$

by Assumption C(i) and C(ii). And lastly for $\tilde{J}_{4,7}$ and $\tilde{J}_{4,8}$ we have

$$\begin{aligned}\left\|\tilde{J}_{4,7}\right\| &= \left\|\tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\lambda'_i W'_F\overline{\hat{W}}_F V_{NL}^{-1}\right\| \\ &\lesssim \left\|\tilde{H}\right\|\left\|\frac{1}{N}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\lambda'_i\right\|\left\|\frac{W'_F\overline{\hat{W}}_F}{T}\right\| \\ &= O_p\left(\gamma_L^{\frac{3}{2}-3d_{F,\max}}\right)\left\|\frac{1}{N}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\lambda'_i\right\| \equiv O_p\left(\gamma_L^{\frac{3}{2}-3d_{F,\max}}\right)\check{J}_{F\varepsilon,j},\end{aligned}$$

and

$$\begin{aligned}\tilde{J}_{4,8} &= \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}W'_{\varepsilon,i}\left(\overline{\tilde{W}_F\tilde{H}}\right)V_{NL}^{-1} \\ &\quad + \tilde{H}'\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}W'_{\varepsilon,i}\overline{\left(\hat{W}_F-\tilde{W}_F\tilde{H}\right)}V_{NL}^{-1} \equiv \tilde{J}_{4,81}+\tilde{J}_{4,82},\end{aligned}$$

where

$$\begin{aligned}\left\|\tilde{J}_{4,81}\right\| &\lesssim \gamma_L^{2-4d_{F,\max}}\left\|\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}W'_{\varepsilon,i}\overline{\tilde{W}}_F\right\| \\ &\leq \gamma_L^{2-4d_{F,\max}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT^2}\sum_{i=1}^{N}\left\|W^*_{\varepsilon,i}\tilde{W}_F\right\|^2\right)^{\frac{1}{2}} \\ &= O_p\left(L^{-\frac{1}{2}}\gamma_L^{\frac{5}{2}-4d_{F,\max}-d_{\varepsilon}}\right)\tilde{J}_{F\varepsilon,j}\end{aligned}$$

by Assumption D(iii). And

$$\left\|\tilde{J}_{4,82}\right\| \lesssim \gamma_L^{1-2d_{F,\max}}\left\|\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F,j}\overline{W}_{\varepsilon,ij}W'_{\varepsilon,i}\overline{\left(\hat{W}_F-\tilde{W}_F\tilde{H}\right)}\right\|$$

$$\leq \gamma_L^{1-2d_{F,\max}} \tilde{J}_{F\varepsilon,j} \left(\frac{1}{NT} \sum_{i=1}^{N} \|W_{\varepsilon,i}\|^2 \right)^{\frac{1}{2}} T^{-\frac{1}{2}} \left\| \hat{W}_F - \tilde{W}_F \tilde{H} \right\|$$
$$\leq O_p \left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} \delta_{NT} \right) \tilde{J}_{F\varepsilon,j}.$$

Then to sum up what we have obtained so far,

$$\left\| \tilde{J}_4 \right\| \leq O_p \left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}} \left\| \tilde{\delta} \right\| + N^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} \right) \tilde{J}_{FF,j}$$
$$+ O_p \left(\gamma_L^{\frac{3}{2}-3d_{F,\max}} \left\| \tilde{\delta} \right\| \right) \tilde{J}_{FX,j} + O_p \left(L^{-\frac{1}{2}} \gamma_L^{\frac{5}{2}-4d_{F,\max}-d_\varepsilon} \right) \tilde{J}_{F\varepsilon,j} + O_p \left(\gamma_L^{\frac{3}{2}-3d_{F,\max}} \right) \check{J}_{F\varepsilon,j}$$

by Assumption E(ii), E(iii) and the convergence rate of $\tilde{\delta}$ given by Lemma C.1.1. And it is easy to see $\left\| \tilde{J}_5 \right\|$ has the same order as $\left\| \tilde{J}_4 \right\|$, and $\left\| \tilde{J}_6 \right\|$ is dominated by $\left\| \tilde{J}_4 \right\|$ and $\left\| \tilde{J}_5 \right\|$ because there are negligible terms with higher order included. The detailed analysis will be very tedious so we omit here. Therefore we can summarize all the terms so far and conclude that

$$\left\| \hat{I}_{F,j} - \tilde{I}_{F,j} \right\| \leq O_p \left(\delta_{NT} \right) \left\| \tilde{I}_{F,j} \right\| + O_p \left(\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_{X,\max}} \left\| \tilde{\delta} \right\| + N^{-\frac{1}{2}} \gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon} \right) \tilde{J}_{FF,j}$$
$$+ O_p \left(\gamma_L^{\frac{3}{2}-3d_{F,\max}} \left\| \tilde{\delta} \right\| \right) \tilde{J}_{FX,j} + O_p \left(L^{-\frac{1}{2}} \gamma_L^{\frac{5}{2}-4d_{F,\max}-d_\varepsilon} \right) \tilde{J}_{F\varepsilon,j}$$
$$+ O_p \left(\gamma_L^{\frac{3}{2}-3d_{F,\max}} \right) \check{J}_{F\varepsilon,j} + I_{E_F,j}$$

with $I_{E_F,j}$ being a smaller order term at each j relative to the others. ■

Bibliography

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81:1203–1227.

Altansukh, G., Becker, R., Bratsiotis, G., and Osborn, D. (2017). What is the globalisation of inflation? *Journal of Economic Dynamics and Control*, 74:1–27.

Andersen, T., Bollerslev, T., Diebold, F., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579–625.

Andrews, D. W. and Sun, Y. (2004). Adaptive local polynomial whittle estimation of long-range dependence. *Econometrica*, 72:569–614.

Andrés, J., Arce, O., Thaler, D., and Thomas, C. (2020). When fiscal consolidation meets private deleveraging. *Review of Economic Dynamics*, 37:214–233.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71:135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77:1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70:191–221.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176:18–29.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.

Banerjee, A. and Urga, G. (2005). Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics*, 129:1–34.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57:33–45.

Beran, J. (2009). On parameter estimation for locally stationarylong-memory processes. *Journal of Statistical Planning and Inference*, 139:900–915.

Boubaker, H. (2018). A generalized arfima model with smooth transition fractional integration parameter. *Journal of Time Series Econometrics*, 10:1–20.

Boutahar, M., Dufrénot, G., and Péguin-Feissolle, A. (2008). A simple fractionally integrated model with a time-varying long memory parameter dt. *Computational Economics*, 31:225–241.

Brillinger, D. (2001). *Time series: data analysis and theory.* Society for Industrial and Applied Mathematics, Philadelphia.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*, chapter Stationary time series, pages 1–41. Springer.

Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167:413–425.

Caporale, G. and Gil-Alana, L. (2007). Nonlinearities and fractional integration in the us unemployment rate. *Oxford Bulletin of Economics and Statistics*, 69:521–544.

Caporale, G. and Gil-Alana, L. (2013). Long memory in us real output per capita. *Empirical Economics*, 44:591–611.

Caporale, G., Gil-Alana, L., and Malmierca, M. (2021). Persistence in the private debt to gdp ratio: Evidence from 43 oecd countries. *Applied Economics*, pages 1–10.

Chambers, M. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review*, pages 1053–1072.

Cheung, Y. (2021). Long memory factor model: on estimation of factor memories. *Journal of Business and Economic Statistics*, pages 1–14.

Christensen, B. and Nielsen, M. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, 133:343–371.

Chung, C. (2002). Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes. *Econometric Theory*, pages 51–78.

Coakley, J., Dollery, J., and Kellard, N. (2011). Long memory and structural breaks in commodity futures markets. *Journal of Futures Markets*, 31:1076–1113.

Corbae, D., Ouliaris, S., and Phillips, P. C. B. (2002). Band spectral regression with trending data. *Econometrica*, 70(3):1067–1109.

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

Dahlhaus, R. (2012). Locally stationary processes. *Handbook of Statistics*, 30:351–413.

Dahlhaus, R. and Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics*, 24:1934–1963.

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

Ding, Z., Granger, C., and Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106.

Doornik, J. and Ooms, M. (2004). Inference and forecasting for arfima models with an application to us and uk inflation. *Studies in Nonlinear Dynamics and Econometrics*, 8:1–23.

Eggertsson, G. and Krugman, P. (2012). Debt, deleveraging, and the liquidity trap: A fisher-minsky-koo approach. *The Quarterly Journal of Economics*, 127:1469–1513.

Ergemen, Y. (2019). System estimation of panel data models under long-range dependence. *Journal of Business and Economic Statistics*, 37:13–26.

Ergemen, Y. and Velasco, C. (2017). Estimation of fractionally integrated panels with fixed effects and cross-section dependence. *Journal of Econometrics*, 196:248–258.

Fan, X., Grama, I., and Liu, Q. (2015). Exponential inequalities for martingales with applications. *Electronic Journal of Probability*, 20:1–22.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192:291–312.

Gil-Alana, L. and Robinson, P. (2001). Testing of seasonal fractional integration in uk and japanese consumption and income. *Journal of Applied Econometrics*, 16:95–114.

Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238.

Haldrup, N. and Vald'es, J. (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics*, 199:1–11.

Hall, P. and Heyde, C. (1980). *Martingale limit theory and its application.* Academic Press.

Hannan, E. J. (1963). Regression for time series with errors of measurement. *Biometrika*, 50(3/4):293–302.

Hannan, E. J. (1970). *Multiple time series*, volume 38. John Wiley & Sons.

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

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business and Economic Statistics*, 13:37–45.

Hidalgo, J. (2003). An alternative bootstrap to moving blocks for time series regression models. *Journal of Econometrics*, 117:369–399.

Hidalgo, J. and Robinson, P. (2002). Adapting to unknown disturbance autocorrelation in regression with long memory. *Econometrica*, 70:1545–1581.

Kang, S., Hernandez, J., and Yoon, S. (2019). Who leads the inflation cycle in europe? inflation cycle and spillover influence among eurozone and non-eurozone economies. *International Economics*, 160:56–71.

Kreiss, J. and Paparoditis, E. (2015). Bootstrapping locally stationary processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 267–290.

Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econometric Theory*, 25:1433–1445.

Kuan, C. and Hsu, C. (1998). Change-point estimation of fractionally integrated processes. *Journal of Time Series Analysis*, 19:693–708.

Lahiri, K. and Liu, F. (2006). Modeling multi-period inflation uncertainty using a panel of density forcasts. *Journal of Applied Econometrics*, 21:1199–1219.

Lazarová, v. (2005). Testing for structural change in regression with long memory processes. *Journal of Econometrics*, 129:329–372.

Lobato, I. (1997). Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis*, 18:137–155.

Lobato, I. (1999). A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics*, 90:129–153.

Lu, Z. and Guegan, D. (2011). Estimation of time-varying long memory parameter using wavelet method. *Communications in Statistics - Simulation and Computation*, 40:596–613.

Luciani, M. and Veredas, D. (2015). Estimating and forecasting large panels of volatilities with approximate dynamic factor models. *Journal of Forecasting*, 34:163–176.

Marinucci, D. and Robinson, P. (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics*, 105:225–247.

Michelacci, C. and Zaffaroni, P. (2000). Fractional beta convergence. *Journal of Monetary Economics*, 45:129–153.

Moon, H. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83:1543–1579.

Mumtaz, H. and Surico, P. (2012). Evolving international inflation dynamics: world and country-specific factors. *Journal of the European Economic Association*, 10:716–734.

Neely, C. and Rapach, D. (2011). International comovements in inflation rates and country characteristics. *Journal of International Money and Finance*, 30:1471–1490.

Nielsen, M. (2005). Semiparametric estimation in time-series regression with long-range dependence. *Journal of Time Series Analysis*, 26:279–304.

Nielsen, M. and Frederiksen, P. (2011). Fully modified narrow-band least squares estimation of weak fractional cointegration. *The Econometrics Journal*, 14:77–120.

Ombao, H. and Ho, R. (2006). Time-dependent frequency domain principal components analysis of multichannel non-stationary signals. *Computational statistics & data analysis*, 50(9):2339–2360.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92:1004–1016.

Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74:967–1012.

Pesaran, M. and Chudik, A. (2014). Aggregation in large dynamic panels. *Journal of Econometrics*, 178:273–285.

Phillips, P. C. B. (1991). Spectral regression for cointegrated time series. In Barnett, W., Powell, J., and Tauchen, G., editors, *Nonparametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press, Cambridge, UK.

Preuß, P. (2012). *Statistical inference for locally stationary long memory processes*. PhD thesis, Ruhr-Universität Bochum.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business and Economic Statistics*, 29:423–438.

Rachinger, H. (2011). Multiple breaks in long memory time series. Universidad Carlos III de Madrid.

Ray, B. and Tsay, R. (2002). Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis*, 23:687–705.

Robinson, P. (1994a). Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields*, 99:443–473.

Robinson, P. (1994b). Semiparametric analysis of long-memory time series. *The Annals of Statistics*, 22:515–539.

Robinson, P. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23:1630–1661.

Robinson, P. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23:1048–1072.

Robinson, P. (1995c). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23:1048–1072.

Robinson, P. and Hidalgo, F. (1997). Time series regression with long-range dependence. *The Annals of Statistics*, 25:77–104.

Roueff, F. and von Sachs, R. (2011). Locally stationary long memory estimation. *Stochastic Processes and their Applications*, 121:813–844.

Shimotsu, K. (2012). Exact local whittle estimation of fractionally cointegrated systems. *Journal of Econometrics*, 169:266–278.

Shimotsu, K. and Phillips, P. (2006). Local whittle estimation of fractional integration and some of its variants. *Journal of Econometrics*, 130:209–233.

Stock, J. and Watson, M. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, 4:351–394.

Stock, J. and Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Vershynin, R. (2018). *High-dimensional probability: an introduction with applications in data science*. Cambridge University Press.

Wang, C., Bauwens, L., and Hsiao, C. (2013). Forecasting a long memory process subject to structural breaks. *Journal of Econometrics*, 177:171–184.

Wang, L. (2019). Time varying long memory parameter estimation for locally stationary long memory processes. *Communications in Statistics-Theory and Methods*, 48:2529–2547.

Wenger, K., Leschinski, C., and Sibbertsen, P. (2018). A simple test on structural change in long-memory time series. *Economics Letters*, 163:90–94.

Whitcher, B. and Jensen, M. (2000). Wavelet estimation of a local long memory parameter. *Exploration Geophysics*, 31:94–103.

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

Xiao, Z. and Phillips, P. C. B. (1998). Higher-order approximations for frequency domain time series regression. *Journal of Econometrics*, 86(2):297–336.