UNIFIED FACTOR MODEL ESTIMATION AND INFERENCE UNDER SHORT AND LONG MEMORY

By

Shuyao Ke, Peter C. B. Phillips and Liangjun Su

October 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2351

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Unified Factor Model Estimation and Inference under Short and Long Memory *

Shuyao Ke[a] Peter C. B. Phillips[b] Liangjun Su[c]

[a]College of Economics, Jinan University, China
[b]Yale University, University of Auckland, Southampton University and Singapore Management University
[c]School of Economics and Management, Tsinghua University, China

October 29, 2022

Abstract

This paper studies a linear panel data model with interactive fixed effects wherein 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. Simulations show that the frequency domain estimator is robust to short memory and outperforms the time domain estimator when long range dependence is present. An empirical illustration of the approach is provided, examining the long-run relationship between stock return and realized volatility.

Key words: Factor error structure, Fourier transform, Frequency domain regression, Interactive fixed effects, Long-range dependence, Principal components.

JEL Classification: C22, C23, C38.

*Preliminary versions of this paper were presented at several seminars. We thank the participants in these presentations, Jia Li, Jun Yu and Sainan Jin for their constructive comments. Phillips acknowledges support from a Lee Kong Chian Fellowship at SMU, the Kelly Fund at the University of Auckland, and the NSF under Grant No. SES 18-50860. Su gratefully acknowledges the support from the National Natural Science Foundation of China under Grant No. 72133002. Address correspondence to: Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, 100084, China. Email: sulj@sem.tsinghua.edu.cn.

1 Introduction

For the past two decades, linear regression panel data models with interactive fixed effects (IFEs) 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 strong cross section dependence via 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. We shall frequently use the abbreviation '(panel) factor model' to represent this general class of models.

For panel 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 linear regression panel data model with IFEs wherein the 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; see, 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 motivate the study of 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 regression panel data model with IFEs with potential long memory regressors, factors and idiosyncratic errors. The starting point of the analysis is standard principal components least squares estimation of 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 nonnegligible 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 the factors contain a constant column or not, 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 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 (DFTs) 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 the 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), Moon and Weidner (2015), and Lu and Su (2016), 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 (2022). 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 employs PCA in the frequency domain to estimate the DFTs of the factors. Similar to our approach but working in a pure factor model, Cheung (2022) seeks to estimate the memory parameters of the latent factors by PCA. Cheung (2022) 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 (2022) 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 2 introduces the factor model with possible long memory in the component variables. Section 3 develops the asymptotics of least squares estimation in the time domain, as in Bai (2009) but allowing for stationary long memory. Section 4 provides the corresponding analysis in the frequency domain. Section 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 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 7 provides an empirical application of our panel frequency domain procedures to investigate the long-run relationship between stock returns and realized volatilities for a monthly panel dataset of 49 industry portfolios. Section 8 concludes. Proofs of the main results and further technical details are provided in the Online Supplement.

The following notations are adopted. For an arbitrary $m \times n$ matrix A, its transpose is denoted by A'; its conjugate and conjugate transpose are denoted $\overline{A}$ and A^* when complex; and its Frobenius norm is $\|A\| = \sqrt{\operatorname{tr}(A'A)}$ if A is real, or $\|A\| = \sqrt{\operatorname{tr}(A^*A)}$ 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. For an $m \times n$ matrix A, $\mathbf{P}_A = A(A'A)^{-1}A'$ and $\mathbf{M}_A = \mathbb{I}_m - \mathbf{P}_A$ when $A'A$ is nonsingular.

2 Model

This paper considers data generated by the linear panel model

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + \varepsilon_{it}, \; i = 1, \ldots, N, \; t = 1, \ldots, T, \tag{2.1}$$

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}. The variables X_{it}, F_t and ε_{it} may be stationary long memory time series with respective memory parameter vectors given by $d_X = (d_{X_1}, \ldots, d_{X_P})'$, $d_F = (d_{F_1}, \ldots, d_{F_R})'$, and d_ε. The memory parameters of both X_{it} and ε_{it} are restricted to be identical across individuals i, so that cross-sectional heterogeneity in memory is induced via cross-sectional heterogeneity in the long memory factors induced by the factor loadings. This sacrifices some generality but facilitates theoretical development.

Among the different ways of defining long memory (e.g., Haldrup and Vald'es, 2017), the linear process approach (Robinson and Hidalgo, 1997) is adopted here. In particular, when d_ε and the elements of d_F and d_X lie in the interval $[0, \frac{1}{2})$, the time series F_t, X_{it} and ε_{it} are assumed to have the following moving average representations:

$$F_t = \mu_F + \sum_{j=0}^{\infty} A_{F,j}\zeta_{F,t-j} \equiv \mu_F + F_t^o, \tag{2.2}$$

$$X_{it} = \mu_{X,i} + \sum_{j=0}^{\infty} A_{X,j}\zeta_{X,i,t-j} \equiv \mu_{X,i} + X_{it}^o \text{ for } i = 1, \ldots, N, \text{ and} \tag{2.3}$$

$$\varepsilon_{it} = \sum_{j=0}^{\infty} A_{\varepsilon,j}\zeta_{\varepsilon,i,t-j}, \tag{2.4}$$

where $A_{F,j}$ and $A_{X,j}$ are $R \times R$ and $P \times P$ coefficient matrices, $A_{\varepsilon,j}$ is a scalar, $\zeta_{F,t}$, $\zeta_{X,i,t}$ and $\zeta_{\varepsilon,i,t}$ are the corresponding innovations, and μ_F and $\mu_{X,i}$ denote the respective means. This specification includes stationary $ARFIMA\,(p, d, q)$ time series as a special case. Differing from the factor innovations $\zeta_{F,t-j}$, the regressor and idiosyncratic error innovations allow for heterogeneity and dependence of X_{it} and ε_{it} across both individuals and time periods, as detailed in Section 3. Following Bai (2009), the least squares (LS) estimators of β and F_t in the time domain are given by the solution to the following nonlinear equations:

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} X_i \right)^{-1} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} Y_i \tag{2.5}$$

where $X_i' = (X_{i1}, \cdots X_{iT})$, $Y_i' = (Y_{i1}, \cdots Y_{iT})$, and

$$\left[\frac{1}{NT} \sum_{i=1}^{N} \left(Y_i - X_i\hat{\beta} \right) \left(Y_i - X_i\hat{\beta} \right)' \right] \hat{F} = \hat{F} V_{NT}, \tag{2.6}$$

under the identification restrictions that $\frac{F'F}{T} = \mathbb{I}_R$ and $\frac{\Lambda'\Lambda}{N}$ is a diagonal matrix. Here $F = (F_1', \ldots, F_T')'$, $\Lambda = (\lambda_1', \ldots, \lambda_N')'$, and V_{NT} is a diagonal matrix that stacks the eigenvalues of the term inside the square brackets in (2.6) in descending order along its primary diagonal. The present study focuses mainly on the asymptotic behavior of $\hat{\beta}$, as developed in the next section.

3 Asymptotic Behavior of Least Squares Estimator

In the following β^0, F_t^0, and F^0 denote the true values of β, F_t, and F, whereas λ_i continues to denote the true value of the factor loadings as it is only implicitly estimated in what follows. 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 used in (Bai, 2009, pp. 1240). Let $Z_i = Z_i(F^0) = (Z_{i1}, ..., Z_{iT})'$, $D_{NT} = D_{NT}(F^0)$, $\zeta_{X,t} = (\zeta_{X,1,t}, \ldots, \zeta_{X,N,t})'$, $\zeta_{\varepsilon,t} = (\zeta_{\varepsilon,1,t}, \ldots, \zeta_{\varepsilon,N,t})'$, and $\gamma_N(s,t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(\varepsilon_{it}\varepsilon_{is})$. For the memory parameters of the regressors and factors, we use the notation $d_{X,\max} = \max_{1\le p\le P} d_{X_p}$, and $d_{F,\max} = \max_{1\le r\le R} d_{F_r}$. Further, let $d_Z = (d_{Z_1}, \ldots, d_{Z_P})'$ be the memory parameter of Z_{it} and set $d_{Z,\max} = \max_{1\le p\le P} d_{Z_p}$. Similarly, set $d_{X,\min} = \min_{1\le p\le P} d_{X_p}$, with corresponding definitions of $d_{F,\min}$ and $d_{Z,\min}$. M is a generic positive constant that may vary across locations.

Note that the nature of $d_{Z,max}$ and $d_{Z,min}$ is not immediately interpretable from the above definition, as in the current studies dealing with long memory variables. This is because the relationship between $Z_i(F)$ and F is nonlinear (in (F_t, X_{it})) in view of the projection geometry. This can complicate the usual memory order relationship since the simple linear relationship theory, wherein the largest long memory parameter dominates in a linear combination, that is used extensively in Cheung (2022), does not necessarily hold. In the present case, most of these complications are avoided by the stationarity and finite moment conditions; but things can be different in nonstationary cases. For instance, in the simple case of scalar $X_t = O_p(1)$ for stationary short memory with $d_X = 0$ and scalar $F_t = O_p(\sqrt{T})$ for a unit root nonstationary factor with long memory $d_F = 1$, we have $Z_t = X_t - \frac{\sum_{s=1}^T X_s F_s}{\sum_{s=1}^T F_s^2} F_t \equiv O_p(1) - \frac{O_p(T)}{O_p(T^2)} \times O_p(\sqrt{T}) \equiv O_p(1) - O_p(\frac{1}{\sqrt{T}}) = O_p(1)$, where the larger memory $(d_F = 1)$ of F_t does not dominate. On the other hand, in the case of unit root scalar $X_t = O_p(\sqrt{T})$ with $d_X = 1$ and scalar short memory factor $F_t = O_p(1)$ with $d_F = 0$, we have $Z_t = X_t - \frac{\sum_{s=1}^T X_s F_s}{\sum_{s=1}^T F_s^2} F_t \equiv O_p(\sqrt{T}) - \frac{O_p(T)}{O_p(T)} \times O_p(1) \equiv O_p(\sqrt{T}) - O_p(1) = O_p(\sqrt{T})$, where now the larger memory $(d_X = 1)$ of X_t does dominate. [1] Therefore we need to consider the memory parameter of Z_i in an explicit manner here so as to avoid the complexity due to the nonlinear structure within.

[1] Similar differences occur in nonstationary long memory cases. Of course, in the present stationary long memory case, there are finite variances and the corresponding law of large numbers apply, so that a normalization condition such as those involved in Assumption B(ii) and (iii) below avoid this issue. But in more general cases, this is an issue that needs to be dealt with, and we leave it to the future extension.

The following assumptions are used in the technical development.

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, the $P \times P$ matrix Π_X and the scalar Π_ε are all nonsingular. Otherwise, assume $A_{a,j}$ is square summable in Frobenius norm.

(ii) $\zeta_{F,t}$, $\zeta_{X,t}$ and $\zeta_{\varepsilon,t}$ satisfy $\mathbb{E}\left(\zeta_{F,t}|\mathcal{F}_{F,t-1}\right) = 0$, $\mathbb{E}\left(\zeta_{X,t}|\mathcal{F}_{X,t-1}\right) = 0$, and $\mathbb{E}\left(\zeta_{\varepsilon,t}|\mathcal{F}_{\varepsilon,t-1}\right) = 0$, 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 analogous notation applies to $\zeta_{X,t}$. We assume that $\zeta_{F,t}$ satisfy

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

and

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

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-, …, p_4-th elements of $\zeta_{F,t}$. Also the analogous condition holds for $\zeta_{X,t}$ and $\zeta_{\varepsilon,t}$. Additionally, $\zeta_{\varepsilon,t}$ satisfies the following eighth-order moment condition

$$\mathbb{E}\left[\zeta_{\varepsilon,t(p_1)}\cdots\zeta_{\varepsilon,t(p_8)} \mid \mathcal{F}_{F,t-1}\right] = \Phi_{4,p_1\ldots p_8} < \infty \tag{3.1}$$

for some absolute constant $\Phi_{4,p_1\ldots p_8}$, and for arbitrary p_1-, …, 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) $\mathbb{E}\left\|X_{it}\right\|^4 \leq M$.

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

(iii) $\mathbb{E}\left\|F_t^0\right\|^4 \leq M$ 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) $\mathbb{E}\left\|\lambda_i\right\|^4 \leq M$ 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) $\mathbb{E}\left(\varepsilon_{it}\right) = 0$ and $\mathbb{E}\left|\varepsilon_{it}\right|^8 \leq M$.

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

$$\frac{1}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij} \leq M, \tag{3.2}$$

and

$$\frac{1}{T^{1+2d_\varepsilon}}\sum_{t,s=1}^{T}\tau_{ts} \leq M,\ \frac{1}{NT^{1+2d_\varepsilon}}\sum_{i,j,t,s=1}\left|\sigma_{ij,ts}\right| \leq M,\ \frac{1}{T^{\max(4d_\varepsilon,1)}}\sum_{t,s=1}^{T}\left|\gamma_N\left(s,t\right)\right|^2 \leq M. \tag{3.3}$$

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

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

Assumption D. (i) Let $\Sigma_{F,ts}=\mathbb{E}\left(F_t^oF_s^{o\prime}\right)$. Assume that $\|\Sigma_{F,ts}\|\leq\tau_{F,ts}$, $\frac{1}{T^{1+2d_{F,\max}}}\sum_{t,s=1}^{T}\tau_{F,ts}\leq$ M,

$$\frac{1}{T^{\max(2d_\varepsilon+2d_{F,\max},1)}}\sum_{t,s=1}^{T}\tau_{ts}\tau_{F,ts}\leq M,\ \text{and}\ \frac{1}{NT^{\max(2d_{F,\max}+2d_\varepsilon,1)}}\sum_{i,j,t,s=1}\left|\sigma_{ij,ts}\right|\tau_{F,ts}\leq M. \tag{3.4}$$

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

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

Assumption E. (i) $\operatorname{plim}_{(N,T)\to\infty}D\left(F^0\right)=D_0$ for some nonrandom positive definite matrix D_0.

(ii) When F^0 does not contain a constant column, $N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\sum_{i=1}^{N}Z_i'\varepsilon_i\overset{d}{\to}\mathcal{N}\left(0,\Sigma\right)$; When there exists a constant column in F^0, $N^{-\frac{1}{2}}T^{\max(d_{Z,\max}+d_\varepsilon,1/2)-1}\sum_{i=1}^{N}Z_i'\varepsilon_i\overset{d}{\to}\mathcal{N}\left(0,\Sigma\right)$. Here, we have $\Sigma=\operatorname{plim}\rho_{NT}^{-2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\sigma_{ij,ts}Z_{it}Z_{js}'$, where $\rho_{NT}=N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}$ in case F^0 contains a constant column and $\rho_{NT}=N^{\frac{1}{2}}T^{1-\max(d_{Z,\max}+d_\varepsilon,1/2)}$ otherwise.

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

Remark 1. Assumption A is a panel data extension of 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 autocovariances of F_t^0, X_{it} and ε_{it} satisfy, as $j\to\infty$

$$\begin{aligned}\Gamma_{F^0}\left(j\right)&=\operatorname{Cov}\left(F_t^0,F_{t-j}^0\right)\sim diag\left(j^{d_F-\frac{1}{2}}\right)C_{F^0}diag\left(j^{d_F-\frac{1}{2}}\right),\\ \Gamma_{X_i}\left(j\right)&=\operatorname{Cov}\left(X_{it},X_{i,t-j}\right)\sim diag\left(j^{d_X-\frac{1}{2}}\right)C_Xdiag\left(j^{d_X-\frac{1}{2}}\right),\end{aligned}$$

and

$$\Gamma_{\varepsilon_i}\left(j\right)=Cov\left(\varepsilon_{it},\varepsilon_{i,t-j}\right)\sim C_\varepsilon j^{2d_\varepsilon-1}, \tag{3.5}$$

for some absolute constant matrices C_{F^0}, 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 $\left\|A_{F^0,j}\right\|^2\leq(1+\delta)^2C\sum_{r=1}^{R}j^{2d_{F_r}-2}$ for some positive constant C when $j\geq K_\delta$, which then implies

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

by Riemann sum approximation if $d_{F,\max}<\frac{1}{2}$. This illustrates how Assumption 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 stationary ARMA model, the rate is usually exponential and thus not nested in this part of Assumption A(i) by simply substituting $d_a = 0$. Therefore in the second part we assume the stationarity of all the variables when some short memory processes are involved. Assumption C below deals with short and long memory uniformly, which is explained in Remark 3. As mentioned above, a widespread alternative approach to long memory process modeling is via $ARFIMA\,(p, d, q)$ specifications that are common in applications of $I\,(d)$ time series. Relative to the fully parametric $ARFIMA\,(p, d, q)$ setting, our formulation involves semiparametric long memory and is free from short-run dynamic specification thereby avoiding potential inconsistent estimation if the parametric autoregressive or moving average components are misspecified. In Assumption A(ii) and A(iii), moment conditions to the eighth order are imposed to assist in the asymptotic development. 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, a condition that rules out dynamic panel models and is also assumed in Bai (2009).

Remark 2. Assumption B is borrowed from the Assumptions A and B in Bai (2009), specifying finite fourth order moments 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 the 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, note that Assumption A(ii) implies the zero expectation, and (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), (3.2) is the standard condition of cross-sectional weak dependence of $\varepsilon_{i,t}$, and the other inequalities specify the serial dependence, as they generalize Assumption C(ii) in Bai (2009) by including long memory. They can be verified by using Theorem 1 of Chung (2002) via a direct application of (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 stationarity, and we can express the bound $\tau_{ts} = \tau_{t-s}$ accordingly. Noting that $\tau_{t-s} = \tau_{s-t}$, we have

$$\frac{1}{T}\sum_{t,s=1}^{T}\tau_{ts} = \frac{1}{T}\sum_{t,s=1}^{T}\tau_{t-s} = \tau_0 + \frac{2}{T}\sum_{t>s}\tau_{t-s} = \frac{2}{T}\sum_{k=1}^{T-1}(T-k)\,\tau_k + O\,(1)\,. \tag{3.6}$$

Let $\gamma_i\,(k)$ be an arbitrary autocovariance function of order k of ε_{it}. By (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

$$\begin{aligned}
\frac{1}{T}\sum_{k=1}^{T-1}(T-k)\tau_k &= \sum_{k=1}^{K_\delta}\left(1-\frac{k}{T}\right)|\gamma(k)| + \sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)|\gamma(k)| \\
&= \sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)|\gamma(k)| + O(1) \\
&\geq C_\varepsilon(1-\delta)\sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)k^{2d_\varepsilon-1} + O(1) \\
&= C_\varepsilon(1-\delta)T^{2d_\varepsilon}\frac{1}{T}\sum_{k=K_\delta}^{T-1}\left(1-\frac{k}{T}\right)\left(\frac{k}{T}\right)^{2d_\varepsilon-1} + O(1) \\
&= C_\varepsilon(1-\delta)T^{2d_\varepsilon}\int_{K_\delta/T}^{1}(1-r)r^{2d_\varepsilon-1}dr\left\{1+O\left(\frac{1}{T}\right)\right\} + O(1), \qquad (3.7)
\end{aligned}$$

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) as long as the cross-sectional correlations among the $\{\varepsilon_{it}\}$ are 'weak enough'. Analogously, for the third inequality in (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

$$T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}(1-r)r^{4d_\varepsilon-2}dr = T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}r^{4d_\varepsilon-2}dr - T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}r^{4d_\varepsilon-1}dr, \qquad (3.8)$$

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

$$\begin{aligned}
T^{4d_\varepsilon-1}\int_{K_\delta/T}^{1}r^{4d_\varepsilon-2}dr &= \int_{K_\delta/T}^{1}(rT)^{4d_\varepsilon-2}d(rT) \equiv \int_{K_\delta}^{T}(r_*)^{4d_\varepsilon-2}d(r_*) \qquad (3.9) \\
&= T^{4d_\varepsilon-1} + O(1) = O(1). \qquad (3.10)
\end{aligned}$$

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). 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 at the end of (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 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 Assumption C(iv) in Bai (2009). We omit them here for brevity.

Remark 4. Assumption D(i) and D(ii) can be verified by using the convergence rate established in Theorem 3 of 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 for Assumption D, take F_t^0, assume $R=1$, and let $\gamma_F(k)$ denote the autocovariance function of F_t^0. Then $\gamma_F(k) \sim C_F k^{2d_F-1}$ for some constant C_F as $k \to \infty$. Following the reasoning in (3.6) and (3.7), we have

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

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 as used in (3.8)-(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 (3.4) in Assumption D(i) holds. Similarly, the second part of (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) and is required for the asymptotic covariance matrix of $\hat{\beta}-\beta^0$. Assumption E(ii) is related to the convergence rate of $\hat{\beta}-\beta^0$. Because of cross-sectional weak dependence of ε_{it}, we can use the Lindeberg-Lévy CLT over i, which requires a uniform boundedness of the second moment of $Z_i'\varepsilon_i$ after certain normalization. For the data generating processes in (2.2)-(2.4), we observe that under the strict exogeneity condition in Assumption A(iv), the temporal dependence of $Z_{it}\varepsilon_{it}$ is dominated by the mean of Z_{it}, denoted by μ_Z when it is nonzero. For a simple illustration, consider $\sum_{t=1}^{T} Z_{it}\varepsilon_{it}$ for arbitrary i. Its mean is zero and its variance-covariance matrix is given by

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

where Z_{it}^o is defined in the same way as X_{it}^o in (2.4). By Assumptions C(ii) and D(ii),

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

Then $\mu_Z\mu_Z'$ dominates in the above summation as along as $\mu_Z \neq 0$, and only the autocovariance structure of ε_{it} is applicable because of its mean-zero nature. Note that by definition, Z_i can be interpreted as the residual of the linear projection of X_i on the column space of F^0, demeaned by a weighted average. So by construction $E(Z_i'F^0) = 0$ holds, and $\mathbb{E}(Z_i) = 0$ if F^0 contains a constant column or if

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

with $\mathbb{E}\left(u_{it} \mid F^0\right) = 0$, is the true data generating process. That is, X_{it} follows a pure factor model with the latent factor given by F_t^0. The latter setting is adopted in some current studies (see Ergemen, 2019, among others) and appears somewhat restrictive but is easy to deal with in practice. In this study we allow both settings described above but only emphasize the former in assumption. If $\mathbb{E}(Z_i) = 0$, the convergence rate is adopted from Theorem 3 in Chung (2002). In pure time series models, we do not obtain asymptotic normality for the OLS estimator when $d_{Z,\max} + d_\varepsilon \geq \frac{1}{2}$; but in panel data models, weak dependence over i and a large number of individual units helps regain asymptotic normality by virtue of cross-section averaging. In addition, a constant column in F^0 indicates the existence of individual fixed effects in our model. In Bai (2009), the presence of individual fixed effects is only a special case where the LS estimator is less efficient by ignoring this feature than the one based on the within-group transformed model. But in our model, individual fixed effects may affect the convergence rate of the LS estimator when long memory exists in the idiosyncratic error term. In the presence of individual fixed effects in our model, it is possible to conduct within-group transformation prior to LS estimation. Readers are referred to the online supplement for discussion of the asymptotic behavior of the LS estimator in the transformed model.

Remark 6. Assumption F is motivated by the notion of fractional cointegration, which generalizes the usual concept of cointegration in the time series literature; see, e.g., Marinucci and Robinson (2001). It also implies that $d_\varepsilon \leq d_{Z,\min}$ by virtue of the 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}, respectively. The following theorem establishes the asymptotic distribution of the LS estimator $\hat{\beta}$.

Theorem 3.1 *Suppose that Assumptions A-F hold and $T/N \to \rho \in (0,\infty)$ as $(N,T) \to \infty$. Then 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(iii), D_0 and Σ are given in Assumption E(i)-(ii), and the bias terms A_{NT} and C_{NT} are each $O_p(1)$ and given by

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

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

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 that the usual convergence rate of the LS estimator $\hat{\beta}$ is slowed by the presence of long memory. In terms of the limit distribution, although asymptotic normality still holds, the bias terms now have orders that are dependent on the long memory parameters, which affect the validity of bias correction based on the usual analytical form and the half-panel jackknife. 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 Theorem 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 $d_{Z,\max}+d_\varepsilon$ exceeds $\frac{1}{2}$. This limit theory substantially complicates the implementation of LS estimation, which is illustrated by Monte Carlo simulations in Section 6. In particular, the traditional analytical bias correction behaves poorly in the presence of long memory, together with poor inference based on the estimator of asymptotic covariance matrix proposed under serial weak dependence. This difficulty in implementation and general poor performance call for an alternative methodology to deal with stationary long memory in panel factor models. The next section develops a new frequency domain least squares (FDLS) approach that extends the use of spectral regression with long memory time series to the panel context.

4 FDLS Estimation and Asymptotic Theory

4.1 Estimation

Transform model (2.1) by taking discrete Fourier transforms (DFTs) for all $i = 1,\dots,N$ over the Fourier frequencies $\gamma_j = \frac{2\pi j}{T}$ for $j = 1,\dots,L$

$$\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}Y_{it}e^{\mathbf{i}t\gamma_j} = \frac{\beta'}{\sqrt{2\pi T}}\sum_{t=1}^{T}X_{it}e^{\mathbf{i}t\gamma_j}+\frac{1}{\sqrt{2\pi T}}\lambda_i'\sum_{t=1}^{T}F_t^0e^{\mathbf{i}t\gamma_j}+\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\varepsilon_{it}e^{\mathbf{i}t\gamma_j}, \tag{4.1}$$

where $\mathbf{i} = \sqrt{-1}$ is the imaginary unit. Recall that use of the frequencies γ_j provides a mean correction in the frequency domain so that the DFT of F_t^0 in (2.2) is

$$\sum_{t=1}^{T} F_t^0 e^{\mathbf{i}t\gamma_j} = \mu_F \sum_{t=1}^{T} e^{\mathbf{i}t\gamma_j} + \sum_{t=1}^{T} F_t^o e^{\mathbf{i}t\gamma_j} = \mu_F \frac{e^{\mathbf{i}\gamma_j}\left(1 - e^{\mathbf{i}T\gamma_j}\right)}{1 - e^{\mathbf{i}\gamma_j}} + \sum_{t=1}^{T} F_t^o e^{\mathbf{i}t\gamma_j} = \sum_{t=1}^{T} F_t^o e^{\mathbf{i}t\gamma_j},$$

since $e^{\mathbf{i}T\gamma_j} = e^{\mathbf{i}2\pi j} = 1$ for all integers $j = 1, \ldots, L$.

For ease of notation, let $W_{Y,ij} = \sum_{t=1}^{T} Y_{it} e^{\mathbf{i}t\gamma_j}$ and similarly define $W_{X,ij}$, $W_{F,j}$ and $W_{\varepsilon,ij}$. 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. Stack $W_{Y,i}$ and $W_{X,i}$ respectively into W_Y and W_X, which are an $N \times L$ matrix and an $N \times L \times P$ tensor. Then (4.1) can be rewritten as

$$W_{Y,ij} = \beta' W_{X,ij} + \lambda_i' W_{F^0,j} + W_{\varepsilon,ij},\ i = 1, \ldots, N,\ j = 1, \ldots, L. \tag{4.2}$$

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

$$\begin{aligned} SSR\left(\beta, W_F, \Lambda\right) &= \frac{1}{NT} \sum_{i=1}^{N} \left(W_{Y,i} - W_{X,i}\beta - W_F\lambda_i\right)^* \left(W_{Y,i} - W_{X,i}\beta - W_F\lambda_i\right) \\ &= \frac{1}{NT} \left\| W_Y - W_X\beta - \Lambda W_F' \right\|^2, \end{aligned} \tag{4.3}$$

subject to the constraints that $\tilde{\Gamma}_F W_F^* W_F \tilde{\Gamma}_F / T = \mathbb{I}_R$ and $\Lambda'\Lambda$ is diagonal, where $\tilde{\Gamma}_F = diag\left\{\gamma_L^{d_{F_r} - 1/2}\right\}$. Here, $W_X\beta = \sum_{p=1}^{P} W_X^p \beta_p$ with β_p and W_X^p corresponding to the p-th element of β and the p-th slice of W_X. Note that $\tilde{\Gamma}_F$ is an $R \times R$ diagonal matrix for normalization over the frequency domain. Such a normalization is justified by the properties of the average periodogram, as considered in the assumptions and remarks below. Let $\tilde{W}_{F,j} = \tilde{\Gamma}_F W_{F,j}$ and $\tilde{\lambda}_i = \tilde{\Gamma}_F^{-1} \lambda_i$ and rewrite the model (4.2) as

$$W_{Y,ij} = \beta' W_{X,ij} + \tilde{\lambda}_i' \tilde{W}_{F,j} + W_{\varepsilon,ij},\ i = 1, \ldots, N,\ j = 1, \ldots, L, \tag{4.4}$$

or in vector-matrix notation

$$W_{Y,i} = W_{X,i}\beta + \tilde{W}_F \tilde{\lambda}_i + W_{\varepsilon,i},\ i = 1, \ldots, N, \tag{4.5}$$

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 $\tilde{W}_F$ span 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 (4.5) by $\mathbf{M}_{\tilde{W}_F}$ to obtain

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

Infeasible FDLS estimation of β obtained by regressing $\mathbf{M}_{\tilde{W}_F} W_{Y,i}$ on $\mathbf{M}_{\tilde{W}_F} W_{X,i}$ yields

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

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

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

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

$$\frac{1}{NT} \mathrm{tr}\left[\left(W_U - \tilde{\Lambda} \tilde{W}_F'\right)^* \left(W_U - \tilde{\Lambda} \tilde{W}_F'\right)\right] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{L} \left| W_{U,ij} - \tilde{\lambda}_i' \tilde{W}_{F,j} \right|^2. \tag{4.6}$$

This objective function is essentially a frequency domain version of that in (Bai, 2009, pp. 1236). Concentrate $\tilde{\Lambda}$ out using

$$\tilde{\Lambda} = W_U \overline{\tilde{W}}_F \left(\tilde{W}_F^* \tilde{W}_F\right)^{-1} = W_U \overline{\tilde{W}}_F / T, \tag{4.7}$$

along with the restriction $\tilde{W}_F^* \tilde{W}_F / T = \mathbb{I}_R$. Then, using (4.7) the objective function in (4.6) becomes

$$\begin{aligned} \mathrm{tr}\left[\left(W_U - \tilde{\Lambda} \tilde{W}_F'\right)^* \left(W_U - \tilde{\Lambda} \tilde{W}_F'\right)\right] &= \mathrm{tr}\left[\left(W_U - W_U \overline{\tilde{W}}_F \tilde{W}_F' / T\right)^* \left(W_U - W_U \overline{\tilde{W}}_F \tilde{W}_F' / T\right)\right] \\ &= \mathrm{tr}\left(W_U^* W_U\right) - \mathrm{tr}\left(\tilde{W}_F' W_U^* W_U \overline{\tilde{W}}_F\right) / T. \end{aligned} \tag{4.8}$$

Minimizing (4.6) is equivalent to maximizing $\mathrm{tr}\left(\tilde{W}_F' W_U^* W_U \overline{\tilde{W}}_F\right)$, yielding a typical principal components analysis (PCA) problem in the 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 in this frequency domain setting and the estimator of $\tilde{W}_F$, denoted by $\hat{W}_F$, is given by the eigenvectors of $W_U^* W_U$ scaled by $\sqrt{T}$ that correspond to its R largest eigenvalues, all of which are real because $W_U^* W_U$ is Hermitian. As in Bai (2009), indeterminacy over rotation for $\tilde{W}_F$ still holds by virtue of the restriction $\tilde{W}_F^* \tilde{W}_F / T = \mathbb{I}_R$. Moreover, this PCA decomposition leads to an estimator of $\tilde{W}_F$, which is normalized column-wise by the matrix $\tilde{\Gamma}_F$. So W_F is not identified here and the same issue holds for $\tilde{\Lambda}$. However, this lack of identifiability is immaterial for estimation of β.

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 to the following set of nonlinear equations:

$$\tilde{\beta} = \left[\sum_{i=1}^{N} \mathrm{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{X,i}\right)\right]^{-1} \sum_{i=1}^{N} \mathrm{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} W_{Y,i}\right), \tag{4.9}$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)^{*}\right]\hat{W}_{F}=\hat{W}_{F}V_{NL}, \tag{4.10}$$

where V_{NL} is the diagonal matrix containing the R largest eigenvalues in decreasing order of the term inside the square brackets of (4.10). Next estimate $\tilde{\Lambda}$ by $\hat{\Lambda}=\hat{W}_{U}\overline{\hat{W}}_{F}/T=(\hat{\lambda}_{1},...,\hat{\lambda}_{N})'$, where $\hat{W}_{U}=\left(\hat{W}_{U,1},\ldots,\hat{W}_{U,N}\right)'$ and $\hat{W}_{U,i}$ denotes the DFT of $\tilde{U}_{i}=U_{i}(\tilde{\beta})$. It is easy to verify that $\hat{\Lambda}'\hat{\Lambda}/N=V_{NL}$. We now develop asymptotic theory for the estimators $\tilde{\beta}$ and $\hat{W}_{F}$.

4.2 Asymptotic properties of the FDLS estimator

To proceed, we start with some notation. Let $\tilde{\Gamma}_{X}=diag(\gamma_{L}^{d_{X_{p}}-\frac{1}{2}})$ and $\tilde{W}_{X,i}=W_{X,i}\tilde{\Gamma}_{X}$ for each i in the same manner as $\tilde{\Gamma}_{F}$ and $\tilde{W}_{F}$ were defined above. Similarly, let $\tilde{\Gamma}_{\varepsilon}=\gamma_{L}^{d_{\varepsilon}-1/2}$ and $\tilde{W}_{\varepsilon_{i}}=W_{\varepsilon,i}\tilde{\Gamma}_{\varepsilon}$. As in the time domain, define

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

where $W_{Z,i}\left(F\right)=\mathbf{M}_{W_{F}}W_{X,i}-\frac{1}{N}\sum_{k=1}^{N}\mathbf{M}_{W_{F}}W_{X,k}a_{ik}$. Let $W_{Z,i}=W_{Z,i}\left(F^{0}\right)$ and $D_{NL}^{\dagger}=D_{NL}^{\dagger}\left(W_{F^{0}}\right)$. Then Z_{it} can be defined in the 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 $f_{\varepsilon,i}\left(\cdot\right)$ denote the marginal spectral density of ε_{it}. We now introduce some extra assumptions that are specified for FDLS estimation.

Assumption A*. (i) Denote the $(P+R+1)$-vector $V_{it}=\left(X_{it}',F_{t}^{0\prime},\varepsilon_{it}\right)'$. For each i, $\{V_{it},t\geq1\}$ is covariance stationary and has spectral density matrix satisfying

$$f_{V,i}\left(\gamma\right)\sim\Gamma\left(\gamma\right)\Upsilon_{i}\Gamma\left(\gamma\right)\ \text{ as }\gamma\rightarrow0^{+}, \tag{4.11}$$

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

$$\Upsilon_{i}=\left(\begin{array}{ccc}\Upsilon_{i,XX}&\Upsilon_{i,XF}&0\\\Upsilon_{i,XF}'&\Upsilon_{FF}&0\\0&0&\Upsilon_{i,\varepsilon\varepsilon}\end{array}\right),$$

in which for each i, the $P\times P$ and $R\times R$ submatrices $\Upsilon_{i,XX}$ and Υ_{FF} are positive definite, and the scalar $\Upsilon_{i,\varepsilon\varepsilon}>0$. $\Gamma\left(\gamma\right)$ is a diagonal matrix given by

$$\Gamma\left(\gamma\right)=diag\left(\gamma^{-d_{X_{1}}},\ldots,\gamma^{-d_{X_{P}}},\gamma^{-d_{F_{1}}},\ldots,\gamma^{-d_{F_{R}}},\gamma^{-d_{\varepsilon}}\right).$$

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

$$\left|f_{V,i,(ab)}-\upsilon_{i,(ab)}\gamma^{-d_{a}-d_{b}}\right|=O\left(\gamma^{\theta-d_{a}-d_{b}}\right)\ \text{ as }\gamma\rightarrow0^{+},$$

for $a, b = 1, \ldots, (P+R+1)$, where $f_{V,i,(ab)}(\gamma)$ denotes the (a,b)th element of $f_{V,i}(\gamma)$ and $\upsilon_{i,(ab)}$ are constants independent of γ.

(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 (2.2)-(2.4). Let $A_V(\gamma) = \sum_{j=0}^{\infty} A_{V,j}e^{\mathbf{i}j\gamma}$. As $\gamma \to 0^+$,

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

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

Assumption B*. (i) Let $\check{\Gamma}_{X,j} = \text{diag}(\gamma_j^{d_{X_p}})$ and assume $\mathbb{E}\left\|\check{\Gamma}_{X,j}W_{X,ij}\right\|^4 \le M$ 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 $\mathcal{W} = \left\{\tilde{W}_F \in \mathbb{C}^{L\times R} : \tilde{W}_F = W_F\tilde{\Gamma}_F,\ \tilde{W}_F^*\tilde{W}_F/T = \mathbb{I}_R\right\}$. Assume $\inf_{\tilde{W}\in\mathcal{W}} D_{NL}^{\dagger}\left(\tilde{W}\right) > 0$.

(iii) Let $\check{\Gamma}_{F,j} = \text{diag}(\gamma_j^{d_{F_r}})$ and assume $\mathbb{E}\left\|\check{\Gamma}_{F,j}W_{F,j}\right\|^4 \le M$ 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) $\mathbb{E}\left\|\gamma_j^{d_\varepsilon}W_{\varepsilon,ij}\right\|^8 \le M$ 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 $\sigma_{ij,kl}^{W,1} = \sqrt{\mathbb{E}\left|W_{\varepsilon,ik}W_{\varepsilon,jl}^*\right|^2}$ and $\sigma_{i,kl}^{W,1} = \mathbb{E}\left|W_{\varepsilon,ik}W_{\varepsilon,il}^*\right|^2$. Assume $\sigma_{ij,kl}^{W,1} \le \gamma_k^{-d_\varepsilon}\gamma_l^{-d_\varepsilon}\overline{\sigma}_{ij}^W$, $\sqrt{\sigma_{i,kl}^{W,1}\sigma_{j,kl}^{W,1}} \le \gamma_k^{-2d_\varepsilon}\gamma_l^{-2d_\varepsilon}\overline{\sigma}_{ij}^W$ for all (k,l) and

$$\frac{1}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij}^W \le M. \tag{4.12}$$

Let $\sigma_{ij,kl}^{W,2} = \left|\mathbb{E}\left(W_{\varepsilon,ik}W_{\varepsilon,jl}^*\right)\right|$. Assume $\left|\sigma_{ij,kl}^{W,2}\right| \le \overline{\sigma}_{kl}^W$ for all (i,j),

$$\frac{\gamma_L^{2d_\varepsilon}}{L^{1+2d_\varepsilon}}\sum_{k,l=1}^{L}\overline{\sigma}_{kl}^W \le M,\ \frac{\gamma_L^{2d_\varepsilon}}{NL^{1+2d_\varepsilon}}\sum_{i,j=1}^{N}\sum_{k,l=1}^{L}\left|\sigma_{ij,kl}^W\right| \le M, \text{ and } \frac{1}{T^{4d_\varepsilon}\log^2 L}\sum_{k,l=1}^{L}\gamma_N^W(k,l)^2 \le M, \tag{4.13}$$

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

(iii) Let $\Omega_i = \check{\Gamma}_\varepsilon\mathbb{E}\left(W_{\varepsilon,i}W_{\varepsilon,i}^*\right)\check{\Gamma}_\varepsilon$, 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), $\mathbb{E}\left|N^{-\frac{1}{2}}\gamma_j^{d_\varepsilon}\gamma_l^{d_\varepsilon}\sum_{i=1}^{N}\left[W_{\varepsilon,ik}W_{\varepsilon,il}^* - E(W_{\varepsilon,ik}W_{\varepsilon,il}^*)\right]\right|^4 \le M$.

(v) $\frac{\gamma_L^{4d_\varepsilon}}{NL^2}\sum_{i,j=1}^{N}\sum_{k,l=1}^{L}|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,ik}^*, W_{\varepsilon,jl}W_{\varepsilon,jl}^*)| \le M$, $\frac{\gamma_L^{4d_\varepsilon}}{N^2L^2}\sum_{i,j,m,n=1}^{N}\sum_{k,l=1}^{L}|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,jk}^*, W_{\varepsilon,ml}W_{\varepsilon,nl}^*)| \le M$, and $\frac{\gamma_L^{4d_\varepsilon}}{NL^2}\sum_{i,j=1}^{N}\sum_{k,l=1}^{L}|\text{cov}(W_{\varepsilon,ik}W_{\varepsilon,il}^*, W_{\varepsilon,jk}W_{\varepsilon,jl}^*)| \le M$.

Assumption D*. (i) Let $\Gamma_Z = diag(\gamma_L^{d_{Z_p}})$ and $\text{plim}_{(N,L)\to\infty}\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}(W_{F^0})\Gamma_Z = D_0^W$ for some matrix $D_0^W > 0$.

(ii) $\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\text{Re}\left(W_{Z,i}^*W_{\varepsilon,i}\right) \xrightarrow{d} \mathcal{N}\left(0,\Sigma_0^W\right)$, where Σ_0^W =plim $\Sigma_{NL}^{\dagger} > 0$, and $\Sigma_{NL}^{\dagger} \equiv \frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^{N}\text{Re}\left(\Gamma_Z W_{Z,i}^*W_{\varepsilon,i}\right)\text{Re}\left(W_{\varepsilon,i}^*W_{Z,i}\Gamma_Z\right)$.

(iii) $\max_{1\leq i\leq N}\mathbb{E}\left\|\frac{\sqrt{L}\gamma_L^{d_\varepsilon-\frac{1}{2}}}{T}W_{\varepsilon,i}^*\zeta_i\right\|^2\leq M$ and $\max_{1\leq i\leq N}\mathbb{E}\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-\frac{1}{2}}}{NT}\sum_{i=1}^N W_{\varepsilon,i}^*\zeta_i\right\|^2\leq M$, for $\zeta_i = \tilde{W}_{X,i}$, $\tilde{W}_F$, and $\tilde{W}_F\lambda_i$.

Assumption E*. (i) $\frac{L}{T}+\frac{1}{L}\to 0$ as $T\to\infty$.

(ii) Let $\overline{d} = \max\{d_{X,\max}, d_{F,\max}, d_{Z,\max}\}$, $\underline{d} = \min\{d_{X,\min}, d_{F,\min}, d_{Z,\min}\}$, and $\Delta d = \overline{d}-\underline{d}$. Assume that $d_\varepsilon<\frac{1}{4}$, $\underline{d}\geq d_\varepsilon$ and $7(\frac{1}{2}-\overline{d})>\frac{1}{2}$.

(iii) $(N^{-\frac{1}{3}}+(L/N)^{1/2}+L^{-1/2})\gamma_L^{-2\Delta d}\to 0$.

Remark 7. Assumption A* imposes standard restrictions on multivariate stationary long memory processes (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 the zero frequency, where a certain power law is satisfied. Under stationary long memory, Assumption A*(i) and A(i) are basically equivalent, but the former assumption in the frequency domain also holds uniformly under short memory. Note that the spectral density matrix is permitted to have heterogeneous constant multipliers across i, which indicates heterogeneous cross-correlation and auto-correlation among different cross-sectional units.

Remark 8. Assumption B* extends the conditions in Assumption B to the frequency domain, consistent with the probability limit for the average (cross-) periodogram shown in Robinson (1994, Theorem 1) and Lobato (1997, Theorem 1) for the univariate and multivariate cases. It is also consistent with Robinson (1995a, Theorem 1 & (3.16)) on the approximation of the spectral density by expectation of the periodogram.

Remark 9. Assumption C* extends Assumption C to the frequency domain, with C*(i) giving the probability limit of averaged periodogram and the fourth order moment of periodogram of the idiosyncratic error as in B*(i) and B*(iii). C*(ii) gives conditions of cross-sectional weak dependence in (4.12) and of serial dependence over frequencies in (4.13) for the DFT of the idiosyncratic errors. The condition in (4.12) is comparable to that in (3.2) in the time domain, and it is imposed to support C*(v). The conditions in (4.13) are based on Robinson (1995b, Theorem 2), which gives the limit of $E(W_{\varepsilon,i}^*W_{\varepsilon,i})$ at the Fourier frequencies. To see this, we use the fact that $|E(W_{\varepsilon,ik}W_{\varepsilon,jl}^*)|\leq \{E(W_{\varepsilon,ik}W_{\varepsilon,jl}^*)^2\}^{1/2}$ and call upon Robinson (1995b, Theorem 1), which in our setting indicates that

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

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

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

$$
\begin{aligned}
&= \frac{M}{(2\pi)^{2d_\varepsilon - 1}} \left(T^{d_\varepsilon + \frac{1}{2}} \frac{1}{T} \sum_{k=1}^{L} \gamma_k^{-\frac{1}{2}} \right)^2 \\
&\approx \frac{M}{(2\pi)^{2d_\varepsilon - 1}} \left(T^{d_\varepsilon + \frac{1}{2}} \gamma_L^{\frac{1}{2}} \left\{ 1 + O\left(\frac{1}{T}\right) \right\} \right)^2 = O\left(\gamma_L^{-2d_\varepsilon} L^{1+2d_\varepsilon} \right),
\end{aligned}
$$

which explains the order in (4.12). By the same reasoning (4.13) is explained by

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

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

Remark 10. Assumption C*(v) gives higher order conditions that likewise mimic those in the time domain. For further exposition denote $\check{W}_{\varepsilon,ij} = \gamma_j^{d_\varepsilon} W_{\varepsilon,ij}$ and taking the first part of C*(v)

$$
\begin{aligned}
\sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \left| \mathrm{cov}(W_{\varepsilon,ik} W_{\varepsilon,ik}^*, W_{\varepsilon,jl} W_{\varepsilon,jl}^*) \right| &= \sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \left| \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \mathrm{cov}(\check{W}_{\varepsilon,ik} \check{W}_{\varepsilon,ik}^*, \check{W}_{\varepsilon,jl} \check{W}_{\varepsilon,jl}^*) \right| \\
&\leq \sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \sqrt{ \mathrm{Var}\left(\check{W}_{\varepsilon,ik} \check{W}_{\varepsilon,ik}^* \right) \mathrm{Var}\left(\check{W}_{\varepsilon,jl} \check{W}_{\varepsilon,jl}^* \right) } \\
&\leq \sum_{i,j=1}^{N} \sum_{k,l=1}^{L} \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \sqrt{ \mathbb{E} \left| \check{W}_{\varepsilon,ik} \check{W}_{\varepsilon,ik}^* \right|^2 \mathbb{E} \left| \check{W}_{\varepsilon,jl} \check{W}_{\varepsilon,jl}^* \right|^2 } \\
&\leq \sum_{k,l=1}^{L} \gamma_k^{-2d_\varepsilon} \gamma_l^{-2d_\varepsilon} \sum_{i,j=1}^{N} \overline{\sigma}_{i,j}^W = O\left(NT^2 \gamma_L^{2-4d_\varepsilon} \right) \\
&= O\left(NL^2 \gamma_L^{-4d_\varepsilon} \right),
\end{aligned}
$$

by (4.12) and Riemann sum approximation.

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

Remark 12. Assumption E*(i) is standard in the literature on narrow-band frequency domain estimation with long memory data, and it is needed as the parametric power law of the spectral density of a long memory process holds locally at the zero frequency. E*(ii) and E*(iii) impose further conditions on the largest memory parameters of the regressors and factors. Clearly, E*(ii) and E*(iii) are satisfied under E*(i) if $\Delta d = 0$ and $N/T \to \rho \in (0, \infty)$. These restrictions could be relaxed if knowledge of the true memory parameters (or consistent estimation of them) of the latent factors is available in advance. With such information W_F can be estimated directly without using $\tilde{\Gamma}_F$ in the normalization.

Under these conditions the asymptotic theory of FDLS estimation of the panel factor model can be established. The first result concerns consistency.

Proposition 4.1 *Suppose Assumptions A-D and A*-B* hold. Then as $(N, T) \to \infty$, we have*
(i) $\tilde{\beta} \xrightarrow{p} \beta$;
(ii) $\tilde{W}^*_{F^0}\hat{W}_F/T$ is asymptotically invertible and $\left\| \mathbf{P}_{\hat{W}_F} - \mathbf{P}_{W_{F^0}} \right\| \xrightarrow{p} 0$.

Proposition 4.1(i) establishes the consistency of $\tilde{\beta}$ and Proposition 4.1(ii) shows that the column space of $\hat{W}_F$ is asymptotically the same as W_{F^0} or $\tilde{W}_{F^0}$. These are used in the subsequent analyses. The limit distribution of $\tilde{\beta}$ is given in the following theorem.

Theorem 4.2 *Suppose Assumptions A, B and A*-E* hold and $N/T \to \rho \in (0, \infty)$ as $(N, T) \to \infty$. Then*

$$\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta} - \beta^0 - A^W_{NT}\right) \xrightarrow{d} \mathcal{N}\left(0, \left(D^W_0\right)^{-1}\Sigma^W_0\left(D^W_0\right)^{-1}\right),$$

where D^W_0 and Σ^W_0 are defined in Assumption D*, the asymptotic bias term A^W_{NT} is given by

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

where $diag\left(\left|W_{\varepsilon,kj}\right|^2\right) = diag\left(\left|W_{\varepsilon,k1}\right|^2, ..., diag\left(\left|W_{\varepsilon,kL}\right|^2\right)\right)$, $\tilde{G} = \left(\frac{\tilde{W}^*_{F^0}\hat{W}_F}{T}\right)^{-1}\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}$, $D^W_{NL} = \gamma_L^{-1}\Gamma_Z D^\dagger_{NL}\left(W_{F^0}\right)\Gamma_Z = O_p(1)$, and $\phi_L = L^{-1}\gamma_L^{2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}$.

The number of frequency ordinates L (or $h \equiv \frac{L}{T}$) can be treated as a frequency domain 'bandwidth' parameter, measuring the width of the region local to the zero frequency that is used in estimation. Because only these frequencies near zero are used, the resulting FDLS estimator $\tilde{\beta}$ converges to the true value at the $(NL)^{-1/2}$-rate, which is slower than the usual parametric $(NT)^{-1/2}$-rate. This reflects the tradeoff between robustness (against long memory) and efficiency. In addition, Theorem 4.2 indicates that we have a single asymptotically nonnegligible bias term that corresponds to the one of order $O(1/T)$ in the time domain LS estimator of Bai (2009). The other bias term of order

$O(1/N)$ in Bai (2009) is asymptotically negligible here because $\sqrt{NL}/N = \sqrt{L/N} = o(1)$ under Assumption E*(i) and the condition that $N/T \to \rho \in (0, \infty)$.

To perform inference the asymptotic bias and variance matrix need to be estimated. But to incorporate the fact that bias order depends on the memory parameters, we require the estimation of all memory parameters. This then motivates the development of an analytic bias correction that does not rely on estimation of the memory parameters. Importantly, use of DFTs in the frequency domain asymptotically removes serial correlation, as indicated by Assumption C*. This leads to weak dependence over both the cross-sectional units and frequencies and it is possible to correct bias analytically.

To achieve this end, we first propose to estimate

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

by replacing W_{F^0}, $\tilde{\lambda}_i$ and $diag\left(|W_{\varepsilon,kj}|^2\right)$,with $\hat{W}_F$, $\hat{\lambda}_i$ and $\hat{\Omega}_k = diag\left(\hat{W}_{\varepsilon,k1}\hat{W}_{\varepsilon,k1}^{*}, \ldots, \hat{W}_{\varepsilon,kL}\hat{W}_{\varepsilon,kL}^{*}\right)$, respectively, where $\hat{W}_{\varepsilon,i} = W_{Y,i} - \hat{W}_{X,i}\tilde{\beta} - \hat{W}_F\hat{\lambda}_i = (\hat{W}_{\varepsilon,i1}, ..., \hat{W}_{\varepsilon,iL})'$. Then $\bar{A}_{NT}^{W}$ can be estimated by

$$\hat{A}_{NT}^{W} = -\frac{1}{NT} \left(\hat{D}_{NL}^{W}\right)^{-1} \sum_{i=1}^{N} \operatorname{Re}\left(W_{X,i}^{*} \mathbf{M}_{\hat{W}_F} \frac{1}{NT} \sum_{k=1}^{N} \hat{\Omega}_k \hat{W}_F \left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1} \hat{\lambda}_i \right), \tag{4.14}$$

where $\hat{D}_{NL}^{W} = \frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(\hat{W}_{Z,i}^{*}\hat{W}_{Z,i}\right)$, $\hat{W}_{Z,i} = W_{X,i}^{*}\mathbf{M}_{\hat{W}_F} - \frac{1}{N}\sum_{k=1}^{N} W_{X,k}^{*}\mathbf{M}_{\hat{W}_F}\hat{a}_{ik}$, and $\hat{a}_{ik} = \hat{\lambda}_i'\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_k$. The bias-corrected estimator of β is simply $\tilde{\beta}^{bc} = \tilde{\beta} - \hat{A}_{NT}^{W}$. For asymptotic variance matrix estimation, we estimate $\left(D_0^W\right)^{-1}\Sigma_0^W\left(D_0^W\right)^{-1}$ by $\left(\hat{D}_{NL}^{W}\right)^{-1}\hat{\Sigma}_{NL}^{W}\left(\hat{D}_{NL}^{W}\right)^{-1}$, where

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

which is justified when the $\left\{\zeta_{\varepsilon,i,t}\right\}$ are independent across i. The following theorem gives the limit distribution of $\tilde{\beta}^{bc}$ and can be used for inference on β.

Theorem 4.3 *Suppose Assumptions A, B and Assumption A*-E* hold and $N/T \to \rho \in (0, \infty)$ as $(N, T) \to \infty$. Suppose that $\left\{\zeta_{\varepsilon,i,t}\right\}$ is independent across i. Then*

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

An important property of the limit theory in Theorem 4.3 is its self-normalization and no memory parameters are involved in the estimation of the asymptotic bias and variance. This property is particularly useful because simulations suggest that inference based on nonparametric plug-in estimators, such as local Whittle estimators of the memory parameters, can perform poorly in finite samples. To appreciate why self normalization works here, observe that

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

as confirmed in the proof of Theorem 4.3.

5 Determination of the Number of Factors

It has so far been assumed that the true number of factors is known, whereas in practice this number has to be determined empirically in most cases. A procedure is therefore needed for consistent estimation of the true number of factors, denoted R^0. For the pure factor model, various methods are available: see Bai and Ng (2002) for an information criterion approach, Onatski (2010) for an 'edge distribution' approach, Ahn and Horenstein (2013) (AH afterwards) for eigenvalue ratio (ER) and growth ratio (GR) approaches, and Jin et al. (2021) for a cross-validation approach. Here we propose to extend AH's ER approach to the FDLS setting. The same intuition as that explained in AH applies here. To proceed, we specify an upper bound $R_{\max} \geq R^0$. The procedure we develop has the following steps based on a modified version of AH.

1. Perform FDLS estimation using $R_{\max}$ factors, as indicated by (4.9) and (4.10) using model (4.4), and correct for bias as in Theorem 4.3. Denote the resulting estimator $\tilde{\beta}_{(R_{\max})}$.

2. Let $\tilde{u}_{it} = Y_{it} - X_{it}'\tilde{\beta}_{(R_{\max})}$. Stack $\{\tilde{u}_{it}\}$ into an $N \times T$ matrix $\tilde{U}$. Then derive the first $(R_{\max}+1)$ largest eigenvalues of $\tilde{U}\tilde{U}'/NT$, denoted by $\tilde{\mu}_{NT,j}$, $j = 1, \ldots, R_{\max}+1$. For $j = 0$, let $\tilde{\mu}_{NT,j} = w(N,T)$ using a mock eigenvalue $w(N,T)$.

3. Define the eigenvalue ratio $ER(j) = \frac{\tilde{\mu}_{NT,j}}{\tilde{\mu}_{NT,j+1}}$. The ER estimate of R^0 is given by $\tilde{R}_{ER} = \max_{0\leq j\leq R_{\max}} ER(j)$.

It is easy to see that $\tilde{u}_{it}$ follows the approximate pure factor model

$$\tilde{u}_{it} = Y_{it} - X_{it}'\tilde{\beta}_{(R_{\max})} = X_{it}'\left(\beta^0 - \tilde{\beta}_{(R_{\max})}\right) + \lambda_i' F_t + \varepsilon_{it}. \tag{5.1}$$

And the time domain LS estimator of β^0 is not used due to the difficulty in correcting its bias using either closed form or jackknife. In terms of the mock eigenvalue, we follow the setup in AH as presented in the following Assumption G. More details of its setup and the choice of $R_{\max}$ will be discussed in the simulation.

To assist in the following analysis it is helpful to introduce some additional notation, as in AH. Let $\psi_k(A)$ be the k-th largest eigenvalue of matrix A. Let $\mu_{NT,k} = \psi_k\left[\left(\Lambda'\Lambda/N\right)\left(F'F/T\right)\right]$ for $k = 1, \ldots, R^0$. Let $\underline{m} = \min(N,T)$, $\bar{m} = \max(N,T)$, and $\lfloor\cdot\rfloor$ the integer part of its real argument. We add the following conditions.

Assumption G. (i) $\operatorname{plim}\mu_{NT,k} = \mu_k$ for some $\mu_k \in (0,\infty)$ and for each $k = 1, \ldots, R^0$. (ii) R^0 is finite. (iii) $w(N,T) \to 0$ and $w(N,T)\underline{m} \to \infty$ as $\underline{m} \to \infty$.

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

(ii) Let E be the $N \times T$ matrix consisting of the elements 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. elements 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 and G_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 the linear combination $W_X \alpha \equiv \sum_{p=1}^{P} \alpha_p W_X^p$ such that W_X^p is an $N \times L$ complex matrix of DFTs of the p-th regressor component, and the $P \times 1$ vector α satisfies $\|\alpha\| = 1$. There exists a constant $b > 0$ such that

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

Assumptions G-I are comparable to Assumptions 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 in the present setup. Assumption J is the frequency domain extension of Assumption NC in Moon and Weidner (2015), which rules out asymptotic collinearity of the regressors. Assumption B in AH imposes conditions on the moments and cross-sectional and serial dependence of the factors, factor loadings and idiosyncratic errors, which are covered by our Assumptions B and C.

With these additional assumptions the following theorem establishes consistency of the factor number estimator $\tilde{R}_{ER}$ and the preliminary FDLS coefficient estimator $\tilde{\beta}_{(R_{\max})}$.

Theorem 5.1 *Let $R_{\max} \geq R^0 \geq 0$ be a fixed integer. Suppose Assumption B, C, G-J and B*, C* and E* hold. Then* $\left\| \tilde{\beta}_{(R_{\max})} - \beta \right\| = O_p\left(\gamma_L^{\frac{1}{2} - d_{X,\max}} \right)$ *and* $\lim_{m\to\infty} \Pr\left(\tilde{R}_{ER} = R^0 \right) = 1$.

Theorem 5.1 focuses attention on the case where $R_{\max}$ is a fixed integer in order to derive the preliminary consistency rate of $\tilde{\beta}_{(R_{\max})}$ in terms of the Frobenius norm. Consistency of $\tilde{R}_{ER}$ indicates that in practice the estimate $\tilde{R}_{ER}$ can be used as a substitute for the true number of factors when N and T are sufficiently large.

6 Monte Carlo Simulations

This section reports the results of Monte Carlo simulations designed to assess the performance of the time and frequency domain estimates of the regression coefficient β^0 and estimates of the factor number R^0.

6.1 Data generating processes (DGPs)

We use two DGPs based on the model

$$Y_{it} = X_{it}'\beta^0 + \lambda_i' F_t^0 + \theta_\varepsilon \varepsilon_{it}, \tag{6.1}$$

where $\beta^0 = (0.1, 0.1)'$, $X_{it} = (X_{1it}, X_{2it})'$, $\lambda_i = (\lambda_{1i}, \lambda_{2i})'$ and $F_t^0 = \left(F_{1t}^0, F_{2t}^0\right)'$. Here, $P = R^0 = 2$, and θ_ε is set to control the signal-to-noise (SN) ratio to be around 4, where the SN ratio is defined by the standard deviation (Std) ratio Std $\left(X_{it}'\beta^0 + \lambda_i' F_t^0\right)$ /Std $(\theta\varepsilon_{it})$. We first generate ε_{it}^o as follows

$$\varepsilon_{it}^o = 0.3\varepsilon_{i,t-1}^o + e_{it}, \tag{6.2}$$

where $e_{it} \sim I(d_e)$ is a fractionally integrated process generated by i.i.d. $\mathcal{N}(0,1)$ innovations. Then we consider cases with and without conditional heteroskedasticity for ε_{it}. For conditional homoskedasticity we set $\varepsilon_{it} = \varepsilon_{it}^o$. For conditional heteroskedasticity we generate ε_{it} as follows

$$\varepsilon_{it} = 0.06\sqrt{\frac{X_{it}'X_{it}}{P}}\varepsilon_{it}^o. \tag{6.3}$$

For the factor process we use the specification, we have $F_t^0 = 0.8\tilde{F}_t^0$, and

$$\tilde{F}_t^0 = 0.4\tilde{F}_{t-1}^0 + e_{f,t}, \tag{6.4}$$

where $e_{f,t} \sim I(d_f)$ is a bivariate fractionally integrated process generated by i.i.d. $\mathcal{N}(0,4)$ innovations with mutually independent components. The regressor X_{pit} is generated by

$$X_{pit} = \sum_{r=1}^{R^0} (\chi_{ri} + \lambda_{ri}) \left(F_{r,t-1}^0 + F_{r,t}^0\right) + X_{p,it}^o, \tag{6.5}$$

with $X_{p,it}^o = 0.5X_{pi,t-1}^o + e_{x,pit}$, $e_{x,pit} \sim I(d_X)$ with i.i.d. $\mathcal{N}(0,9)$ innovations same as above. This regressor DGP, adopted from Moon and Weidner (2015), ensures the correlation with factors. $\chi_{ri} \sim \mathcal{N}(1,1)$ and $\lambda_{ri} \sim \mathcal{N}(1,1)$ for each $r = 1, 2$. The innovations $e_{f,t}$, e_{it}, $e_{x,1it}$, $e_{x,2it}$, and λ_{1i}, λ_{2i}, χ_{1i} and χ_{2i} are all mutually independent.

Combining the two error term cases we have two DGPs designated below as DGP1 and DGP2: DGP1 combines (6.2) and (6.4) with conditional homoskedasticity and DGP2 combines (6.3) and (6.4) with conditional heteroskedasticity. In terms of sample size, we consider $N, T = 100$, 200 and the case $N = 50$ and $T = 260$ which fits the dataset in our following application. Various values of the memory parameters are considered for d_e, d_f and d_X such that the memories among the elements of R^0 factors and P regressors are homogeneous. We use 300 replications for each case below and results are presented for the second element of all bivariate vectors.

6.2 Time and frequency domain least squares estimation

First, we examine the finite sample performance of the time and frequency domain LS estimators $\hat{\beta}$ and $\tilde{\beta}$. For evaluation, we focus on the bias (BIAS), standard deviation (STD) and root mean square error (RMSE). For inference, we consider the coverage probability (COVP) of the 95% confidence intervals based on the asymptotic normal critical values, which for the time domain estimator are calculated using the asymptotic covariance estimator proposed by Bai (2009) as if there were no long memory; and in the frequency domain the critical values are derived using the self-normalized inference scheme given in Theorem 4.3.

The following combinations of memory parameters are considered, covering various cases of stationary short and long memory: (1) $d_f = d_e = d_X = 0$, (2) $d_f = d_X = 0.2$ and $d_e = 0.1$, (3) $d_f = d_X = 0.3$ and $d_e = 0.2$, (4) $d_f = d_X = 0.4$ and $d_e = 0.2$. Case (1) involves only short memory and cases (2)-(4) involve stationary long memory time series. We further consider the case $d_e = 0.2$, $d_f = d_X = 0.6$ which partly covers the setup in our empirical application. We include such memory cases to illustrate the performance of our estimator when the regressors and factors are nonstationary long memory processes, although our present theory does cover such cases yet.

The results for time domain estimation are presented in the left panel of Table 1 and Table 2 for each DGP respectively, where bias corrections are conducted using the analytic formulae of Bai (2009) to obtain $\hat{\beta}^{bc}$. Clearly, the time domain LS estimator, after bias correction, performs well in terms of coverage probability only under short memory or weak long memory as in case (2). This estimator typically has mostly prominent downward bias when the joint memory is strong enough. In addition, the convergence rate contaminated by memory parameters adversely affects the accuracy of inference from the results of empirical coverage probabilities.

For FDLS estimation DFTs of (6.1) are calculated over the Fourier frequencies $\gamma_j = \frac{2\pi j}{T}$ for $j = 1, \ldots, L$, with bandwidth $L = \lfloor T^{0.5} \rfloor$ and $L = \lfloor T^{0.6} \rfloor$ respectively. Note that these two settings of bandwidth, together with memory parameters given by (1)-(4) above, satisfy our Assumption E*(ii) and (iii). The right two panels in Table 1 and 2 present the results for the bias-corrected FDLS estimator $\tilde{\beta}^{bc}$ studied in Theorem 4.3. From these results it is evident that under short memory, the FDLS estimator performs almost as well as the time domain LS estimator: both estimators exhibit good bias control and coverage probabilities. This occurs even though the FDLS estimator is asymptotically less efficient due to its use only of frequencies close to zero, and the bias is better controlled for the time domain LS estimator. In contrast, when stationary long memory exists, especially when joint long memory is stronger the FDLS estimator significantly outperforms the time domain LS estimator, prominently correcting for bias in the right direction and most importantly showing good coverage probability in all cases. There is no systematic gain by setting a wider bandwidth from $L = \lfloor T^{0.5} \rfloor$ to $L = \lfloor T^{0.6} \rfloor$ except for efficiency when all long memory parameters are stationary. Nevertheless, when both factors and regressors are nonstationary, the bias correction in FDLS does not outperform the time domain estimator in DGP2 with conditional heteroskedasticity and $L = \lfloor T^{0.5} \rfloor$. Note that no prominent over-coverage is observed so our self-normalized inference is valid, as one may worry about how larger the standard error of FDLS relative to the time domain one will solely explain the increase of coverage probabilities.

6.3 Estimation of R^0

Now we study the finite sample performance of the factor number estimator $\tilde{R}_{ER}$ for both DGPs when the true number of factors R^0 equal to either 2 or 4. In both cases we set $R_{\max} = 8$, as indicated by our a priori information about the maximal value of R^0 being 4. Although not presented here, we find a larger choice of $R_{\max}$ does not change the results prominently. Moreover, the mock eigenvalue is set to be $w(N,T) = \sum_{k=1}^{m} \tilde{\mu}_{NT,k} / \log(\underline{m})$ as suggested by Ahn and Horenstein (2013). For the sole purpose of comparison we also consider the information criterion (IC) proposed by Bai (2009), viz.,

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

where $V\left(r, \hat{F}^r\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2(r)$, $\hat{\varepsilon}_{it}(r)$ is the residual in the time domain LS estimation when r factors are used, and $\hat{F}^r$ is the associated estimator of the factor matrix. For both $\tilde{R}_{ER}$ and the IC estimator, we report the average (Mean), the median (Median), the ratios of correct estimation (RCE), over-estimation (ROE) and under-estimation (RDE) of the true number R^0 in Tables 3–4 for DGPs 1 and 2 respectively, under the bandwidth $L = \lfloor T^{0.6} \rfloor$. As evident from the results in Tables 3–4, in DGP1 both estimators can consistently estimate the R^0 under short memory and weakly long memory, but $\tilde{R}_{ER}$ significantly outperforms the IC estimator when joint memory is strong in DGP1 and under all setups in DGP2. This illustrates the robustness of ER estimator compared to the IC one. One shortcoming for the $\tilde{R}_{ER}$ estimator is that it may suffer from underestimation in the nonstationary case in DGP 2 when the sample size is small, but its performance quickly improves as the sample size increases.

7 Empirical Application

The methodology in this study is applied to re-investigate the relationship between stock returns and realized volatilities, which has been an essential theme of asset pricing literature. The pioneering work by French et al. (1987), Campbell and Hentschel (1992) and Duffee (1995) have established the negative relationship between the aggregate levels of stock returns and realized volatilities using value-weighted market portfolio. Such results are confirmed by the follow-up studies such as Dutt and Humphery-Jenner (2013) using global aggregate level data and by Ang et al. (2006, 2009) using firm level data. Conversely, Duffee (1995) and Grullon et al. (2012) find the evidence of positive relationship between the returns and volatilities in firm level instead. To explain such a contradiction, Grullon et al. (2012) suggest there might be some aggregate market conditions that affect both the market returns and volatilities simultaneously, which motivates us to take such conditions into consideration by modeling them through the use of common factors.

Moreover, the nature of long memory of the realized volatility, as aforementioned, has been well documented by the current literature (Andersen et al., 2003; Christensen and Nielsen, 2007, among others). In a recent study, Liu (2022) tries to decompose the volatility into two components, the more persistent long-run component modeled by unit root and the less persistent short-run component. And the negative relationship between returns and volatilities is found to be significant for the long-run component. This study emphasizes how the persistent part of volatility, which can be modeled by long memory as well, can heavily explain the return-volatility relation, and therefore it is important to consider and handle the long memory as indicated by our theory. In this application we estimate a parsimonious factor model using the FDLS approach. Specifically, we consider the regression of excess returns of individual industry portfolios, Rex_{it}, on the contemporaneous and lagged values of industry-level volatility, VOL_{it} and the first-difference of the log of averaged firm sizes $Size_{i,t}$, which are known in month t, viz.,

$$Rex_{it} = \beta_1 VOL_{it} + \beta_2 \Delta \log Size_{it} + \lambda_i' F_t^0 + \varepsilon_{it}, \tag{7.1}$$

where λ_i and F_t^0 are factor loadings and factors that are of R-dimension, the regressand $Rex_{it} = R_{it} - R_{ft}$, where R_{it} is the value-weighted average return of stocks within industry i at month t, and R_{ft} is the risk-free rate at the same time. The regressor $\Delta \log Size_{it} = \log Size_{it} - \log Size_{i,t-1}$ is a control variable that measures the change of market capitalization of industry i; and the monthly volatility VOL_{it} is calculated using the daily returns following Ergemen and Velasco (2017), as

$$VOL_{it} = \left(\sum_{s \in t}^{N_t} Rex_{is}^2 \right)^{\frac{1}{2}}, \ t = 1, \ldots, T,$$

where Rex_{is} is the excess return of industry i at day s, and N_t is the number of trading days in month t. Different from some current studies, we do not include control variables like book-to-market ratio explicitly, but they can still be well-controlled by the factors and factor loadings. Our dataset is adopted from Kenneth French's Data Library for 49 industries in the U.S. listed companies, with time spanning from 2000 to 2021, giving a balanced panel with $N = 49$ and $T = 263$.

Table 5 presents the descriptive statistics of the data for all industries and all three variables involved, with local Whittle estimates (denoted by $\hat{d}$) of the memory parameters for each variable with bandwidth $L = \lfloor T^{0.6} \rfloor$. We can see that for most of industries, Rex_{it} and $\Delta \log Size_{it}$ have short memory only as the local Whittle estimator is not significant at up to 10%-level, while for some industries these two variables turn out to be either antipersistent ($\hat{d} < 0$) or stationary long memory. Meanwhile VOL_{it} exhibits a very strong evidence of being long memory, with some industries lying in the slightly nonstationary range with memory parameter around 0.6.[2]

[2]While the theoretic framework in the present paper does not include cases where memory parameters in the

Table 6 presents the bias-corrected FDLS estimates of β_1 and β_2 together with their standard errors. To estimate the number of factors, we set $R_{\max} = 8$ and obtain the estimate $\tilde{R}_{ER} = 4$. In fact, this estimate is obtained for all $R_{\max} \in [4, 8]$. As a robustness check, we also consider the other values of R for the estimation by varying it from 3 to 8 in Table 6. Obviously, the results indicate that the realized volatility has a negative effect on the stock returns at the industry level, which is consistent with the early findings in the literature. Our inclusion of factor influences and a long memory structure also support the argument of Liu (2022) and respond to the concern about missing variable issues in Grullon et al. (2012).

8 Conclusion

This paper studies estimation and inference in a linear panel data model with interactive fixed effects by allowing for stationary long memory behavior in the regressors, the factors and the idiosyncratic errors. In this broad context, the usual time domain least squares and principal components approach produces estimates with asymptotic biases that are difficult to correct in practice, as well as poor inference due the complication caused by long memory. The alternative frequency domain least squares estimation developed here takes advantage of the spectral behavior at low frequencies associated with the possible presence of long memory time series. The new approach has favorable asymptotic properties and simulations show that the procedure is well behaved in finite samples.

Several extensions of the present framework and results are possible. First, using the same framework, memory parameters of the factors and idiosyncratic errors can be obtained using local Whittle estimation with the estimated factors and residuals, for which we can develop asymptotic theory as in the current long memory literature. Second, as in Ergemen (2019), it is possible to consider a panel regression where the memory parameters for the regressors and the errors are heterogeneous across individuals or/and the regression coefficients are heterogeneous. Third, it is worth extending the asymptotic theory to allow for stationary antipersistent data with $d \in (-\frac{1}{2}, 0)$. Finally, extensions of the approach and the limit theory to include nonstationary or nearly nonstationary long memory data within the same panel factor model framework would substantially widen the compass of potential applications of this frequency domain approach. These challenges are left for future research.

regressors and errors are heterogeneous across individuals, we still employ the FDLS method to fit this empirical model. Indeed, even though we conjecture that heterogeneity among the memory parameters does not affect consistency of the FDLS estimator of the regression coefficients, it will still affect the rates of convergence, the variance matrix and inference. Hence, what is emphasized here is the long memory nature of the data as a whole and how it is accommodated using our new frequency domain methodology.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81:1203–1227.

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.

Ang, A., Hodrick, R., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Ang, A., Hodrick, R., Xing, Y., and Zhang, X. (2009). High idiosyncratic volatility and low returns: international and further u.s. evidence. *Journal of Financial Economics*, 91(1):1–23.

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.

Brillinger, D. (2001). *Time series: data analysis and theory*. Society for Industrial and Applied Mathematics, Philadelphia.

Campbell, J. and Hentschel, L. (1992). No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318.

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.

Chambers, M. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review*, 39:1053–1072.

Cheung, Y. L. (2022). Long memory factor model: On estimation of factor memories. *Journal of Business & Economic Statistics*, 40(2):756–769.

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.

Christensen, B. and Nielsen, M. (2007). The effect of long memory in volatility on stock market fluctuations. *The Review of Economics and Statistics*.

Chung, C. (2002). Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes. *Econometric Theory*, 18(1):51–78.

Corbae, D., Ouliaris, S., and Phillips, P. C. B. (2002). Band spectral regression with trending data. *Econometrica*, 70(3):1067–1109.

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.

Duffee, G. (1995). Stock returns and volatility: a firm-level analysis. *Journal of Financial Economics*, 37(3):399–420.

Dutt, T. and Humphery-Jenner, M. (2013). 2013. *Journal of Banking and Finance*, 37(3):999–1017.

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.

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.

French, K., Schwert, G., and Stambaugh, R. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29.

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.

Grullon, G., Lyandres, E., and Zhdanov, A. (2012). Real options, volatility, and stock returns. *The Journal of Finance*, 67(4):1499–1537.

Haldrup, N. and Vald'es, J. (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics*, 199:1–11.

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.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business and Economic Statistics*, 13:37–45.

Jin, S., Miao, K., and Su, L. (2021). On factor models with random missing: EM estimation, inference, and cross validation. *Journal of Econometrics*, 222(1):745–777.

Liu, Y. (2022). The short-run and long-run components of idiosyncratic volatility and stock returns. *Management Science*, 68(2):1573–1589.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1313.

Lobato, I. (1997). Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis*, 18:137–155.

Lu, X. and Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190(1):148–175.

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.

Nielsen, M. (2005). Semiparametric estimation in time-series regression with long-range dependence. *Journal of Time Series Analysis*, 26:279–304.

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.

Robinson, P. (1994). Semiparametric analysis of long-memory time series. *The Annals of Statistics*, pages 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. and Hidalgo, F. (1997). Time series regression with long-range dependence. *The Annals of Statistics*, 25:77–104.

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.

Xiao, Z. and Phillips, P. C. B. (1998). Higher-order approximations for frequency domain time series regression. *Journal of Econometrics*, 86(2):297–336.

Table 1: LS estimates over the time and frequency domains (DGP1)

	$\hat{\beta}^{bc}$ (bias-corrected time domain LS estimate by Bai (2009))				$\tilde{\beta}^{bc}$ (bias-corrected FDLS estimate)							
					$L = \lfloor T^{0.5} \rfloor$				$L = \lfloor T^{0.6} \rfloor$			
	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)
$d_e = 0,\ d_f = 0,\ d_X = 0\ (\theta_\varepsilon = 1.2)$												
$N, T = 100$	-0.037	0.339	0.341	90%	0.027	0.785	0.784	93.67%	-0.031	0.639	0.639	96%
$N, T = 200$	-0.000	0.172	0.171	94%	-0.001	0.439	0.439	96.67%	0.011	0.342	0.342	95.67%
$N = 50, T = 260$	-0.027	0.312	0.313	92.34%	-0.041	0.626	0.626	96.5%	-0.041	0.625	0.626	92.34%
$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2\ (\theta_\varepsilon = 1.4)$												
$N, T = 100$	-0.062	0.433	0.436	83.34%	0.012	0.811	0.810	95%	-0.043	0.687	0.688	95.67%
$N, T = 200$	-0.010	0.219	0.219	87%	-0.003	0.441	0.440	95.67%	-0.002	0.358	0.358	94.67%
$N = 50, T = 260$	-0.054	0.385	0.388	83.34%	-0.084	0.842	0.846	95%	-0.062	0.652	0.654	95%
$d_e = 0.2,\ d_f = 0.3,\ d_X = 0.3\ (\theta_\varepsilon = 1.5)$												
$N, T = 100$	-0.072	0.562	0.565	73.67%	0.029	0.898	0.897	95.34%	-0.040	0.770	0.770	96.34%
$N, T = 200$	-0.015	0.283	0.283	74.67%	0.008	0.488	0.487	95%	-0.003	0.403	0.403	94%
$N = 50, T = 260$	-0.076	0.505	0.510	70.34%	0.018	0.852	0.852	95.67%	-0.008	0.680	0.680	95%
$d_e = 0.2,\ d_f = 0.4,\ d_X = 0.4\ (\theta_\varepsilon = 2)$												
$N, T = 100$	-0.091	0.699	0.704	70.67%	0.036	1.035	1.034	94.34%	-0.043	0.903	0.903	97%
$N, T = 200$	-0.020	0.347	0.347	71%	0.019	0.543	0.543	95%	0.008	0.466	0.465	94.67%
$N = 50, T = 260$	0.100	0.634	0.640	64.67%	-0.069	1.010	1.010	94.67%	-0.074	0.836	0.837	94.34%
$d_e = 0.2,\ d_f = 0.6,\ d_X = 0.6\ (\theta_\varepsilon = 3.1)$												
$N, T = 100$	-0.069	0.812	0.814	65.67%	0.011	1.154	1.153	95.34%	-0.062	1.000	1.000	93.67%
$N, T = 200$	-0.056	0.397	0.401	61%	0.004	0.571	0.570	94%	-0.020	0.520	0.519	93%
$N = 50, T = 260$	-0.017	0.678	0.677	61.67%	-0.008	0.955	0.953	95%	-0.012	0.850	0.849	95.34%

Note: All the results of BIAS, STD and RMSE are multiplied by 100 for ease of illustration. We present the selection of θ_ε under each setup.

Table 2: LS estimates over the time and frequency domains (DGP2)

	$\hat{\beta}^{bc}$ (bias-corrected time domain LS estimate by Bai (2009))				$\tilde{\beta}^{bc}$ (bias-corrected FDLS estimate)							
					$L = \lfloor T^{0.5} \rfloor$				$L = \lfloor T^{0.6} \rfloor$			
	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)	BIAS	STD	RMSE	COVP(95%)
$d_e = 0,\ d_f = 0,\ d_X = 0\ (\theta_\varepsilon = 1.6)$												
$N, T = 100$	-0.035	0.401	0.402	90%	0.019	0.823	0.822	93%	-0.056	0.663	0.664	93.67%
$N, T = 200$	-0.027	0.185	0.187	92.34%	-0.018	0.426	0.426	96.5%	-0.024	0.339	0.339	95.34%
$N = 50, T = 260$	0.025	0.341	0.341	92%	-0.101	0.829	0.834	96.5%	-0.167	0.641	0.661	95.67%
$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2\ (\theta_\varepsilon = 1.6)$												
$N, T = 100$	-0.082	0.436	0.429	82.34%	0.038	0.774	0.774	95.67%	-0.046	0.648	0.649	94.34%
$N, T = 200$	-0.022	0.222	0.223	86.67%	-0.010	0.409	0.408	95%	-0.026	0.344	0.345	95.67%
$N = 50, T = 260$	-0.065	0.395	0.400	85.67%	-0.138	0.772	0.783	94%	-0.149	0.630	0.647	94.67%
$d_e = 0.2,\ d_f = 0.3,\ d_X = 0.3\ (\theta_\varepsilon = 1.4)$												
$N, T = 100$	-0.121	0.588	0.600	77.34%	0.053	0.956	0.956	93.67%	-0.043	0.819	0.819	94%
$N, T = 200$	-0.086	0.302	0.314	75%	0.029	0.517	0.517	94.67%	-0.032	0.434	0.434	96.34%
$N = 50, T = 260$	-0.027	0.539	0.539	79.67%	-0.027	0.926	0.925	93.67%	-0.014	0.769	0.766	95%
$d_e = 0.2,\ d_f = 0.4,\ d_X = 0.4\ (\theta_\varepsilon = 1.2)$												
$N, T = 100$	-0.117	0.557	0.569	74%	0.034	0.869	0.869	93%	-0.045	0.759	0.759	94.67%
$N, T = 200$	-0.084	0.291	0.303	70%	0.025	0.463	0.463	94.67%	-0.024	0.398	0.398	95.34%
$N = 50, T = 260$	-0.041	0.539	0.539	73%	0.009	0.848	0.846	92.34%	-0.017	0.726	0.725	94.34%
$d_e = 0.2,\ d_f = 0.6,\ d_X = 0.6\ (\theta_\varepsilon = 1.1)$												
$N, T = 100$	-0.144	0.689	0.703	64%	0.287	1.438	1.464	90%	0.112	1.109	1.113	91%
$N, T = 200$	-0.052	0.357	0.360	66%	0.087	0.555	0.561	93.67%	0.044	0.520	0.521	92.67%
$N = 50, T = 260$	-0.063	0.668	0.671	64.34%	-0.070	1.015	1.016	94.67%	-0.032	0.901	0.901	93.67%

Note: All the results of BIAS, STD and RMSE are multiplied by 100 for ease of illustration. We present the selection of θ_ε under each setup.

Table 3: Estimation of R^0 (DGP1)

		$R^0 = 2$					$R^0 = 4$				
Method		Mean	Median	RCE	ROE	RDE	Mean	Median	RCE	ROE	RDE
IC		$d_e = 0,\ d_f = 0,\ d_X = 0$ ($\theta_\varepsilon = 1.5$)									
	$N, T = 100$	2	2	1	0	0	4	4	1	0	0
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
		$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$ ($\theta_\varepsilon = 1.4$)									
	$N, T = 100$	2	2	1	0	0	4	4	1	0	0
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
		$d_e = 0.2,\ d_f = 0.3,\ d_X = 0.3$ ($\theta_\varepsilon = 1.8$)									
	$N, T = 100$	2.810	3	0.227	0.773	0	4.657	5	0.390	0.610	0
	$N, T = 200$	2.890	3	0.153	0.847	0	4.713	5	0.303	0.697	0
	$N = 50, T = 260$	2.010	2	0.990	0.010	0	4.003	4	0.997	0.003	0
		$d_e = 0.2,\ d_f = 0.4,\ d_X = 0.4$ ($\theta_\varepsilon = 2.5$)									
	$N, T = 100$	2.683	3	0.347	0.653	0	4.440	4	0.580	0.420	0
	$N, T = 200$	2.760	3	0.263	0.737	0	4.463	4	0.543	0.457	0
	$N = 50, T = 260$	2.010	2	0.990	0.010	0	4.003	4	0.997	0.003	0
		$d_e = 0.2,\ d_f = 0.6,\ d_X = 0.6$ ($\theta_\varepsilon = 3.1$)									
	$N, T = 100$	2.450	2	0.580	0.420	0	4.223	4	0.780	0.220	0
	$N, T = 200$	2.507	2	0.510	0.490	0	4.130	4	0.870	0.130	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
ER		$d_e = 0,\ d_f = 0,\ d_X = 0$									
	$N, T = 100$	2	2	1	0	0	4	4	1	0	0
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
		$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$									
	$N, T = 100$	2	2	1	0	0	4	4	1	0	0
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
		$d_e = 0.2,\ d_f = 0.3,\ d_X = 0.3$									
	$N, T = 100$	2	2	1	0	0	4	4	1	0	0
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
		$d_e = 0.2,\ d_f = 0.4,\ d_X = 0.4$									
	$N, T = 100$	2	2	1	0	0	3.990	4	0.997	0	0.003
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
		$d_e = 0.2,\ d_f = 0.6,\ d_X = 0.6$									
	$N, T = 100$	1.977	2	0.977	0	0.023	3.950	4	0.980	0	0.020
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	3.990	4	0.997	0	0.003

Note: In the parentheses we set the θ_ε's under $R^0 = 4$, so as to maintain the proper SN ratio.

Table 4: Estimation of R^0 (DGP2)

		$R^0 = 2$					$R^0 = 4$				
Method		Mean	Median	RCE	ROE	RDE	Mean	Median	RCE	ROE	RDE
IC		$d_e = 0,\ d_f = 0,\ d_X = 0$ ($\theta_\varepsilon = 1.6$)									
	$N, T = 100$	4.877	5	0.050	0.950	0	5.743	6	0.153	0.847	0
	$N, T = 200$	4.107	4	0.093	0.907	0	5.120	5	0.307	0.693	0
	$N = 50, T = 260$	2.953	3	0.383	0.617	0	4.240	4	0.790	0.210	0
		$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$ ($\theta_\varepsilon = 1.4$)									
	$N, T = 100$	5.970	6	0.023	0.977	0	6.377	6	0.053	0.947	0
	$N, T = 200$	5.553	5.500	0.010	0.990	0	6.150	6	0.080	0.920	0
	$N = 50, T = 260$	3.497	3	0.247	0.753	0	4.490	4	0.600	0.400	0
		$d_e = 0.2,\ d_f = 0.3,\ d_X = 0.3$ ($\theta_\varepsilon = 1.5$)									
	$N, T = 100$	7.323	8	0	1	0	7.433	8	0.007	0.993	0
	$N, T = 200$	7.610	8	0	1	0	7.563	8	0	1	0
	$N = 50, T = 260$	4.767	5	0.020	0.980	0	5.207	5	0.277	0.723	0
		$d_e = 0.2,\ d_f = 0.4,\ d_X = 0.4$ ($\theta_\varepsilon = 1.2$)									
	$N, T = 100$	7.220	8	0	1	0	7.367	8	0.007	0.993	0
	$N, T = 200$	7.520	8	0	1	0	7.470	8	0	1	0
	$N = 50, T = 260$	4.673	4.500	0.023	0.977	0	5.227	5	0.277	0.723	0
		$d_e = 0.2,\ d_f = 0.6,\ d_X = 0.6$ ($\theta_\varepsilon = 1.1$)									
	$N, T = 100$	7.040	7.500	0	1	0	7.383	8	0.007	0.993	0
	$N, T = 200$	7.200	8	0	1	0	7.323	8	0.003	0.997	0
	$N = 50, T = 260$	4.897	5	0.067	0.933	0	5.510	5	0.227	0.773	0
ER		$d_e = 0,\ d_f = 0,\ d_X = 0$									
	$N, T = 100$	2	2	1	0	0	4	4	1	0	0
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	4	4	1	0	0
		$d_e = 0.1,\ d_f = 0.2,\ d_X = 0.2$									
	$N, T = 100$	2	2	1	0	0	4	4	1	0	0
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	3.997	4	0.997	0	0.003
		$d_e = 0.2,\ d_f = 0.3,\ d_X = 0.3$									
	$N, T = 100$	1.993	2	0.993	0	0.007	3.900	4	0.947	0.003	0.050
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	3.997	4	0.997	0	0.003
		$d_e = 0.2,\ d_f = 0.4,\ d_X = 0.4$									
	$N, T = 100$	1.997	2	0.997	0	0.003	3.893	4	0.950	0	0.050
	$N, T = 200$	2	2	1	0	0	4	4	1	0	0
	$N = 50, T = 260$	2	2	1	0	0	3.987	4	0.993	0	0.007
		$d_e = 0.2,\ d_f = 0.6,\ d_X = 0.6$									
	$N, T = 100$	1.950	2	0.950	0	0.050	3.143	4	0.630	0.007	0.363
	$N, T = 200$	1.987	2	0.987	0	0.013	3.820	4	0.927	0	0.073
	$N = 50, T = 260$	1.970	2	0.970	0	0.030	3.557	4	0.820	0	0.180

Note: In the parentheses we set the θ_ε's under $R^0 = 4$, so as to maintain the proper SN ratio.

Table 5: Descriptive statistics and memory parameter estimation in the application

	Rex					VOL					$\Delta Size$				
Industry	Mean	STD	Max	Min	$\hat{d}$	Mean	STD.	Max	Min	$\hat{d}$	Mean	STD	Max	Min	$\hat{d}$
Agric	0.834	6.333	20.97	-18.20	-0.053	6.625	4.242	37.59	1.12	0.626_a	0.012	0.226	1.642	-2.625	0.035
Food	0.683	3.871	17.56	-11.12	0.041	4.079	2.234	21.94	1.58	0.270_a	0.006	0.056	0.249	-0.427	-0.017
Soda	0.989	6.082	28.48	-22.46	-0.107	5.548	3.259	23.49	1.64	0.392_a	0.013	0.132	1.810	-0.253	0.011
Beer	0.710	4.151	11.37	-14.76	-0.072	4.442	2.733	28.67	1.51	0.400_a	0.005	0.078	0.403	-0.822	-0.033
Smoke	1.341	6.465	32.38	-22.18	0.147	5.865	3.374	26.86	2.01	0.520_a	0.007	0.077	0.281	-0.454	0.075
Toys	0.708	7.030	22.77	-23.40	-0.072	6.950	3.396	31.38	2.52	0.539_a	0.010	0.083	0.601	-0.268	-0.007
Fun	1.110	8.011	39.30	-31.60	-0.013	7.785	4.581	35.84	2.25	0.505_a	0.007	0.123	0.739	-0.764	0.332_b
Books	0.340	6.207	30.73	-25.27	0.123	5.953	3.795	26.85	1.99	0.452_a	-0.001	0.099	0.460	-0.699	0.031
Hshld	0.595	3.984	11.40	-14.73	-0.167_c	4.336	2.687	25.29	1.85	0.313_a	0.007	0.045	0.112	-0.160	-0.067
Clths	1.147	6.331	24.91	-21.69	-0.082	6.471	3.489	30.03	2.30	0.411_a	0.013	0.065	0.224	-0.209	-0.133
Hlth	0.864	6.143	20.57	-19.55	-0.080	5.703	3.166	30.06	1.80	0.308_a	0.009	0.065	0.212	-0.229	0.024
MedEq	0.962	4.671	13.92	-19.24	-0.030	5.063	2.702	25.18	1.96	0.263_a	0.009	0.056	0.173	-0.341	-0.099
Drugs	0.661	4.236	13.14	-11.10	0.213_b	4.860	2.565	22.13	1.71	0.280_a	0.001	0.053	0.126	-0.302	-0.120
Chems	0.816	6.085	19.05	-21.06	-0.159_c	6.197	3.616	29.25	2.34	0.436_a	0.004	0.070	0.303	-0.353	-0.212_b
Rubbr	0.954	6.097	31.94	-20.84	-0.074	5.634	2.922	23.25	1.87	0.496_a	0.011	0.077	0.417	-0.259	-0.051
Txtls	0.904	9.191	58.92	-36.09	-0.047	7.802	4.894	39.17	2.60	0.548_a	0.008	0.161	1.435	-1.444	-0.016
BldMt	0.909	6.792	34.40	-31.89	-0.097	6.272	3.813	34.41	2.37	0.415_a	0.005	0.078	0.294	-0.408	-0.155_c
Cnstr	1.039	7.314	21.86	-32.15	-0.019	8.131	4.434	36.53	3.18	0.421_a	0.010	0.086	0.198	-0.420	0.033
Steel	0.550	9.260	26.24	-32.99	-0.006	9.229	4.841	40.91	3.73	0.500_a	0.004	0.099	0.233	-0.400	0.015
FabPr	0.740	8.397	30.37	-32.63	-0.103	8.253	4.378	41.60	2.45	0.394_a	0.010	0.138	1.143	-0.824	-0.072
Mach	1.018	7.012	23.02	-29.83	-0.189_b	6.908	3.992	31.77	2.56	0.463_a	0.009	0.073	0.202	-0.346	-0.184_c
ElcEq	0.699	6.709	22.87	-24.78	-0.053	6.672	3.849	33.10	2.32	0.430_a	-0.003	0.137	0.216	-1.823	0.086
Autos	1.041	9.467	49.56	-36.50	0.108	7.767	4.478	32.88	2.57	0.588_a	0.009	0.095	0.420	-0.438	0.019
Aero	0.903	6.922	32.50	-36	0.004	6.482	4.169	40.86	2.25	0.415_a	0.005	0.075	0.279	-0.444	-0.063
Ships	1.336	7.832	29.15	-22.61	-0.067	7.227	3.397	27.80	2.53	0.432_a	0.005	0.142	0.256	-1.855	0.068
Guns	1.274	5.858	18.40	-21.77	0.010	5.996	3.337	30.21	2.56	0.464_a	0.009	0.066	0.216	-0.245	0.176_c
Gold	0.825	10.329	33.90	-33.61	-0.039	10.644	4.512	43.98	4.27	0.360_a	0.008	0.128	0.785	-0.408	-0.001
Mines	1.159	8.752	26.95	-34.83	-0.073	8.775	4.825	41.96	3.49	0.424_a	0.009	0.095	0.236	-0.432	-0.009
Coal	0.848	13.676	43.54	-40.85	0.130	13.017	6.864	50.99	3.91	0.590_a	-0.008	0.198	0.404	-2.134	-0.134_a
Oil	0.665	7.010	32.92	-34.81	0.063	7.096	4.271	39.32	2.07	0.433_a	0.005	0.073	0.293	-0.411	0.021
Util	0.749	4.178	11.23	-13.14	0.130	4.614	3.101	30.23	1.64	0.290_a	0.008	0.042	0.147	-0.141	0.151
Telcm	0.224	5.228	21.20	-16.30	0.270_a	5.187	3.357	27.64	1.76	0.508_a	0.004	0.055	0.191	-0.175	0.093
PerSv	0.563	5.889	18.59	-26.22	-0.077	6.021	3.022	28.30	2.52	0.334_a	0.009	0.067	0.331	-0.305	-0.075
BusSv	0.704	5.189	18.16	-24.03	0.082	5.260	3.216	31.57	1.69	0.361_a	0.008	0.055	0.167	-0.262	0.090_a
Hardw	0.599	8.046	24.94	-33.88	0.091	7.472	4.702	29.46	2.27	0.619_a	0.003	0.098	0.231	-0.835	0.092
Softw	0.672	6.650	23.83	-22.83	0.110	6.346	3.938	27.12	1.45	0.510_a	0.008	0.071	0.230	-0.235	0.065_a
Chips	0.811	8.240	26.84	-32.62	0.109	7.736	4.718	30.16	2.27	0.622_a	0.009	0.089	0.445	-0.383	0.135_a
LabEq	1.007	6.546	20.49	-23.13	0.014	6.286	3.541	24.74	1.94	0.503_a	0.012	0.072	0.240	-0.292	-0.061_a
Paper	0.565	5.136	23.10	-18.31	-0.026	5.166	2.865	24.31	1.85	0.426_a	0.005	0.056	0.208	-0.258	-0.008
Boxes	0.887	5.861	18.06	-19.64	-0.171_c	6.027	3.142	27.76	2.30	0.359_a	0.009	0.066	0.175	-0.320	-0.205_b
Trans	0.847	5.526	17.06	-16.57	-0.081	5.968	3.152	28.90	2.40	0.465_a	0.010	0.065	0.433	-0.190	-0.107
Whlsl	0.730	5.014	15.93	-21.09	0.065	5.056	2.894	28.33	1.81	0.307_a	0.008	0.052	0.140	-0.246	-0.027
Rtail	0.795	4.863	18.64	-14.92	0.036	5.296	2.921	21.46	1.94	0.447_a	0.009	0.051	0.186	-0.153	-0.105
Meals	0.100	4.873	18.64	-22.47	-0.042	5.152	3.006	31.68	1.79	0.351_a	0.011	0.055	0.171	-0.253	-0.015
Banks	0.588	6.283	19.71	-27.23	0.014	6.866	5.450	35.41	1.67	0.707_a	0.005	0.072	0.183	-0.502	-0.062
Insur	0.792	5.492	22.40	-26.86	0.157_c	5.504	4.163	33.03	1.66	0.527_a	0.007	0.058	0.182	-0.298	0.100
RlEst	0.863	8.275	66.01	-37.59	-0.017	6.536	5.124	37.76	2.07	0.628_a	0.011	0.106	0.756	-0.511	0.059
Fin	0.789	7.082	19.51	-26.20	0.014	7.488	5.034	33.62	1.40	0.614_a	0.011	0.100	1.107	-0.306	-0.009_a
Other	0.355	5.819	21	-22.35	0.064	5.666	3.789	26.60	1.65	0.598_a	0.008	0.200	2.468	-1.571	-0.055

Note: The subscripts a, b and c of $\hat{d}$ indicate the significance at 1%, 5% and 10% nominal levels respectively. And for the maximal and minimal values, we round to only two decimal places for simplicity.

Table 6: Estimation results in the application

R	3	4	5	6	7	8
$\tilde{\beta}_1^{bc}$	-0.095	-0.116	-0.134	-0.117	-0.129	-0.127
s.e.($\tilde{\beta}_1^{bc}$)	(0.034)	(0.033)	(0.030)	(0.026)	(0.021)	(0.021)
$\tilde{\beta}_2^{bc}$	0.732	0.916	0.982	0.982	1.027	1.135
s.e.($\tilde{\beta}_2^{bc}$)	(0.110)	(0.105)	(0.088)	(0.085)	(0.078)	(0.068)

Online Supplement for "Unified Factor Model Estimation and Inference under Short and Long Memory"

Shuyao Ke[a], Peter C. B. Phillips[b] and Liangjun Su[c]

[a]College of Economics, Jinan University, China

[c]Yale University, University of Auckland, Southampton University, and Singapore Management University

[c]School of Economics and Management, Tsinghua University, China

This supplement has three parts. Section A contains the proofs of the main results in Sections 3–5 of the paper. Section B provides proofs of the lemmas stated in Section A. Section C presents some results for the time domain least squares estimation based on the within-group transformed equation.

A Proofs of the Main Results in Sections 3–5

This section proves the main results in Sections 3–5. These proofs call upon some technical lemmas that are proved in Section B. Throughout the present proofs, we use C to denote a constant that may vary according to position and $A \lesssim B$ to denote $A/B = O_p(1)$.

A.1 Proof of Theorem 3.1

Tor prove Theorem 3.1, we need the following four lemmas.

Lemma A.1 *Suppose Assumptions A–E and the other conditions in Theorem 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*

$$\frac{1}{T}\left\|\hat{F} - F^0H\right\|^2 = O_p\left(\left\|\hat{\beta} - \beta^0\right\|^2 + \delta_{NT}^{-2}\right).$$

Lemma A.2 *Suppose Assumptions A–E and the other conditions in Theorem 3.1 hold. Then 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 A.3 *Suppose Assumptions A–E and the other conditions in Theorem 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* $\Omega_k = \mathbb{E}\left(\varepsilon_k\varepsilon_k'\right)$ *for* $k = 1,\ldots,N$.

Lemma A.4 *Suppose Assumptions A–E and the other conditions in Theorem 3.1 hold. Then*

$$\frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{\hat{F}}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{\hat{F}}\right]\varepsilon_i$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{F^0}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{F^0}\right]\varepsilon_i+C_{NT}^{o}+o_p\left(\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\delta_{NT}^{-1}+N^{-\frac{1}{2}}T^{d_\varepsilon-1}\right),$$

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 3.1. By the definition of $\hat{\beta}$ in (2.5), we have

$$\left(\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}X_i\right)\left(\hat{\beta}-\beta^0\right)=\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}F^0\lambda_i+\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\varepsilon_i. \tag{A.1}$$

For the first term on the right hand side (r.h.s.) of the last equation, we notice that $\mathbf{M}_{\hat{F}}F^0=\mathbf{M}_{\hat{F}}(F^0-\hat{F}H^{-1})$, where H is as defined in Lemma A.1: $H=\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right)V_{NT}^{-1}\equiv G^{-1}V_{NT}^{-1}$. 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 continues to hold under long range dependence. 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 equation (2.6)

$$\begin{aligned}\hat{F}V_{NT} &= \frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\hat{\delta}'X_i'\hat{F}-\frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\lambda_i'F^{0\prime}\hat{F}-\frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\varepsilon_i'\hat{F}\\ &+\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i\hat{\delta}'X_i'\hat{F}-\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\hat{\delta}'X_i'\hat{F}+\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i\varepsilon_i'\hat{F}\\ &+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\lambda_i'F^{0\prime}\hat{F}+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\hat{F}+\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i\lambda_i'F^{0\prime}\hat{F}\\ &\equiv I_1+\cdots+I_9,\end{aligned} \tag{A.2}$$

so that $\hat{F}V_{NT}G-F^0=(I_1+\cdots+I_8)\,G$ and

$$\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}F^0\lambda_i=-\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left[I_1+\cdots+I_8\right]G\lambda_i\equiv J_1+\cdots+J_8.$$

The order of the J_ℓ elements are derived by the same reasoning as in the proof of Lemma A.1. Starting with J_1, we have

$$\|J_1\|=\left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}X_k\hat{\delta}\hat{\delta}'X_k'\hat{F}G\lambda_i\right\|$$

$$\leq \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|X_i\|\|\lambda_i\|\frac{1}{NT}\sum_{k=1}^{N}\|X_k\|^2\left\|\hat{\delta}\right\|^2\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\|G\|$$

$$\lesssim \left\|\hat{\delta}\right\|^2\left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right)^{\frac{1}{2}}\frac{1}{NT}\sum_{k=1}^{N}\|X_k\|^2 = O_p\left(\left\|\hat{\delta}\right\|^2\right),$$

where we use the fact that $\left\|\mathbf{M}_{\hat{F}}\right\|_{\text{sp}} = 1$, $\frac{1}{\sqrt{T}}\left\|\hat{F}\right\| = \sqrt{R}$, and Assumption B(i) and B(iv). For J_3, we have

$$J_3 = \frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\frac{X_i'\mathbf{M}_{\hat{F}}X_k}{T}\right)\left(\frac{\varepsilon_k'F^0H}{T}\right)G\lambda_i\hat{\delta}$$

$$+\frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{X_i'\mathbf{M}_{\hat{F}}X_k}{T}\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}G\lambda_i\hat{\delta} \equiv J_{31}+J_{32}.$$

Note that

$$\|J_{31}\| = \left\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\frac{X_i'\mathbf{M}_{\hat{F}}X_k}{T}\right)\left(\frac{\varepsilon_k'F^0H}{T}\right)G\lambda_i\hat{\delta}\right\|$$

$$\lesssim \left\|\hat{\delta}\right\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|X_i\|\|\lambda_i\|\left\|\frac{\varepsilon_k'F^0H}{T}\right\|\frac{1}{N\sqrt{T}}\sum_{k=1}^{N}\|X_k\|$$

$$\leq \left\|\hat{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\left[\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'F^0H}{T}\right\|^2\right]^{\frac{1}{2}}$$

$$\lesssim \left\|\hat{\delta}\right\|\left[\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'F^0H}{T}\right\|^2\right]^{\frac{1}{2}} \equiv \left\|\hat{\delta}\right\|\bar{J}_{31},$$

by Assumption B(i) and B(iv). By the arguments used to show (B.2) in the proof of Lemma A.2 and Assumption D(i), we find 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 Lemma A.1. Then $\|J_3\| = o_p\left(\left\|\hat{\delta}\right\|\right)$. The same approach shows 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

$$\|J_4\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left(\frac{1}{NT}\sum_{k=1}^{N}F^0\lambda_k\hat{\delta}'X_k'\hat{F}\right)G\lambda_i\right\|$$

$$= \frac{1}{N^2T^2}\left\|\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\left(F^0-\hat{F}H^{-1}\right)\sum_{k=1}^{N}\lambda_k\hat{\delta}'X_k'\hat{F}G\lambda_i\right\|$$

$$\lesssim \left\{\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right\}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right\}^{1/2}\frac{\left\|F^0-\hat{F}H^{-1}\right\|}{\sqrt{T}}\left\|\hat{\delta}\right\| = o_p\left(\left\|\hat{\delta}\right\|\right),$$

by Lemma A.1. For J_6, we have

$$\begin{aligned}
\|J_6\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \left(F^0 - \hat{F}H^{-1}\right) \left(\frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{\varepsilon_k' \hat{F}}{T} \right) G\lambda_i \right\| \\
&\lesssim \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} \|X_i\| \|\lambda_i\| \left\| \frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{\varepsilon_k' \hat{F}}{T} \right\| \frac{1}{\sqrt{T}} \left\| F^0 - \hat{F}H^{-1} \right\| \\
&\lesssim \left\| \frac{1}{N} \sum_{k=1}^{N} \lambda_k \frac{\varepsilon_k' \hat{F}}{T} \right\| \frac{1}{\sqrt{T}} \left\| F^0 - \hat{F}H^{-1} \right\| \\
&= O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left(1 + \left\|\hat{\delta}\right\|\right)\right) O_p\left(\left\|\hat{\delta}\right\| + \delta_{NT}^{-1}\right) \\
&= o_p\left(\left\|\hat{\delta}\right\|\right) + O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \delta_{NT}^{-1}\right),
\end{aligned}$$

by Lemmas A.1 and A.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 employ the following decomposition:

$$\begin{aligned}
J_8 &= -\frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \varepsilon_k \varepsilon_k' \hat{F} G \lambda_i \\
&= -\frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \Omega_k \hat{F} G \lambda_i - \frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \left(\varepsilon_k \varepsilon_k' - \Omega_k\right) F^0 H G \lambda_i \\
&- \frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \left(\varepsilon_k \varepsilon_k' - \Omega_k\right) \left(\hat{F} - F^0 H\right) G \lambda_i \equiv J_{81} + J_{82} + J_{83}.
\end{aligned}$$

By Lemma A.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

$$\frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} F^0 \lambda_i = J_2 + J_7 + A_{NT}^o + o_p\left(\left\|\hat{\delta}\right\|\right) + O_p\left(N^{-1} T^{d_\varepsilon - \frac{1}{2}} + N^{-\frac{1}{2}} T^{3d_\varepsilon - \frac{3}{2}}\right). \tag{A.3}$$

Combining (A.1) and (A.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 implies that

$$\begin{aligned}
&\left[D_{NT}\left(\hat{F}\right) + o_p(1)\right] \left(\hat{\beta} - \beta^0\right) \\
&= \frac{1}{NT} \sum_{i=1}^{N} \left(X_i' \mathbf{M}_{\hat{F}} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k' \mathbf{M}_{\hat{F}} \right) \varepsilon_i + A_{NT}^o + O_p\left(N^{-\frac{1}{2}} T^{2d_\varepsilon - 1} + N^{-\frac{1}{2}} T^{3d_\varepsilon - \frac{3}{2}}\right).
\end{aligned} \tag{A.4}$$

Then, by Lemma A.4 and the conditions $\frac{T}{N} \to \rho > 0$ and $d_{Z,\max} > d_\varepsilon$,

$$
\begin{aligned}
&\left[D\left(\hat{F}\right) + o_p(1)\right] \rho_{NT}\left(\hat{\beta} - \beta^0\right) \\
&= \frac{\rho_{NT}}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' \mathbf{M}_{\hat{F}}\right]\varepsilon_i + \rho_{NT} A_{NT}^o + \rho_{NT} O_p\left(N^{-\frac{1}{2}} T^{2d_\varepsilon - 1} + N^{-\frac{1}{2}} T^{3d_\varepsilon - \frac{3}{2}}\right) \\
&= \frac{\rho_{NT}}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' \mathbf{M}_{F^0}\right]\varepsilon_i + \rho_{NT}\left(A_{NT}^o + C_{NT}^o\right) + \rho_{NT} O_p\left(N^{-\frac{1}{2}} T^{2d_\varepsilon - 1} + N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \delta_{NT}^{-1}\right), \\
&= \frac{\rho_{NT}}{NT}\sum_{i=1}^{N} Z_i'\varepsilon_i + \rho_{NT}\left(A_{NT}^o + C_{NT}^o\right) + o_p(1),
\end{aligned}
$$

recalling 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 A.1, we deduce 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

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

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 3.1. ■

A.2 Proofs of the results in Section 4

To prove Proposition 4.1, we use the following lemma.

Lemma A.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 4.1. (i) The proof follows closely that of Proposition 1 in Bai (2009, pp. 1264). Let $\delta = \beta - \beta^0$. By definition, the FDLS estimator $\left(\tilde{\beta}, \hat{W}_F\right)$ solves the following concentrated minimization problem

$$
\left(\tilde{\beta}, \hat{W}_F\right) = \arg\min_{\beta \in \mathbb{R}^P, \tilde{W}_F \in \mathcal{W}} S_{NT}\left(\beta, \tilde{W}_F\right),
$$

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 (4.3) and by (4.4)

$$\begin{aligned}SSR(\beta,W_F,\Lambda) &= \sum_{i=1}^{N}(W_{Y,i}-W_{X,i}\beta-W_F\lambda_i)^*(W_{Y,i}-W_{X,i}\beta-W_F\lambda_i)\\ &= \sum_{i=1}^{N}\left(W_{Y,i}-W_{X,i}\beta-\tilde{W}_F\tilde{\lambda}_i\right)^*\left(W_{Y,i}-W_{X,i}\beta-\tilde{W}_F\tilde{\lambda}_i\right).\end{aligned}$$

Let $W_{U,i}=W_{U,i}(\beta)=W_{Y,i}-W_{X,i}\beta$. As in (4.7), we concentrate out $\tilde{\lambda}_i$ 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 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

$$\begin{aligned}S_{NT}\left(\beta,\tilde{W}_F\right) &= \tilde{S}_{NT}\left(\beta,\tilde{W}_F\right)+\delta'\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^*\mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i}+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^*\mathbf{M}_{\tilde{W}_F}W_{X,i}\delta\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\lambda_i'W_{F^0}^*\mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i}^*+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^*\mathbf{M}_{\tilde{W}_F}W_{F^0}\lambda_i\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}^*\left(\mathbf{P}_{\tilde{W}_F}-\mathbf{P}_{\tilde{W}_{F^0}}\right)W_{\varepsilon,i},\end{aligned}$$

where

$$\begin{aligned}\tilde{S}_{NT}\left(\beta,\tilde{W}_F\right) &= \delta'\left(\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^*\mathbf{M}_{\tilde{W}_F}W_{X,i}\right)\delta+\operatorname{tr}\left[\left(\frac{W_{F^0}^*\mathbf{M}_{\tilde{W}_F}W_{F^0}}{T}\right)\left(\frac{\Lambda'\Lambda}{N}\right)\right]\\ &\quad+\delta'\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^*\mathbf{M}_{\tilde{W}_F}W_{F^0}\lambda_i+\frac{1}{NT}\sum_{i=1}^{N}\lambda_i'W_{F^0}^*\mathbf{M}_{\tilde{W}_F}W_{X,i}\delta.\end{aligned}$$

By Lemma A.5, $S_{NT}\left(\beta,\tilde{W}_F\right)=\tilde{S}_{NT}\left(\beta,\tilde{W}_F\right)+o_p(1)$, uniformly over $\beta\in\mathbb{R}^P$ and $\tilde{W}_F\in\mathcal{W}$, so that it is sufficient to focus on the approximate objective function $\tilde{S}_{NT}\left(\beta,\tilde{W}_F\right)$. 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. Let

$$A = \frac{1}{NT}\sum_{i=1}^{N} W_{X,i}^* \mathbf{M}_{\tilde{W}_F} W_{X,i},\ B = \frac{1}{T}\left(\frac{\Lambda'\Lambda}{N} \otimes \mathbb{I}_L\right),\ C = \frac{1}{NT}\sum_{i=1}^{N}\left(\lambda_i \otimes \mathbf{M}_{\tilde{W}_F} W_{X,i}\right),$$

and $\eta = \text{vec}(\mathbf{M}_{\tilde{W}_F} W_{F^0})$, where $\text{vec}(\cdot)$ vectorizes by stacking columns. 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 $\delta = \beta - \beta^0 \neq 0$ 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, p. 1265), we conclude that $\tilde{\beta}$ is a consistent estimator for β.

(ii) Given consistency of $\tilde{\beta}$, the proof follows exactly as in Proposition 1 in Bai (2009, p. 1265). ■

To prove Theorem 4.2 the following lemmas are employed.

Lemma A.6 *Suppose Assumption A, B, and A*-E*, and the other conditions of Theorem 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 A.7 *Suppose Assumptions A, B, and A*-E*, and the other conditions of Theorem 4.2 hold. Then*

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i\left(\frac{W_{\varepsilon,i}^*\hat{W}_F}{T}\right) = O_p\left(\delta_{W,NL}\left\|\tilde{\beta} - \beta^0\right\| + N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right)$$

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 A.8 *Suppose Assumptions A, B, and A*-E*, and suppose the other conditions of Theorem 4.2 hold. Let* $\tilde{J}_8 = -\frac{1}{NT}\sum_{i=1}^{N} Re\Big(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N} W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\Big)$ *and*

$$A_{NT} = -\frac{1}{NT}\sum_{i=1}^{N} Re\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N} diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\check{G}\lambda_i\right),$$

where diag $\left(|W_{\varepsilon,kj}|^2\right)$ *denotes the diagonal matrix formed from* $|W_{\varepsilon,kj}|^2$, $j = 1,\ldots,L$. *Then*

$$\tilde{J}_8 = A_{NT} + O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$$
$$+ O_p\left(\left(T^{2d_\varepsilon-1}\gamma_L^{d_{F,\min}-d_{X,\max}} + N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon+\left(d_{F,\min}-d_{X,\max}\right)}\right)\left(\delta_{W1,NT}\left(\tilde{\beta}-\beta^0\right) + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)$$

where $A_{NT} = O_p\left(\frac{1}{L}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$.

Lemma A.9 *Suppose Assumptions A, B, and A*-E*, and suppose the other conditions of Theorem 4.2 hold. Recall that* $W_{V,i} = \frac{1}{N}\sum_{k=1}^N a_{ik}W_{X,k}$. *Then*

$$\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N Re\left[\left(W_{X,i}^* - W_{V,i}^*\right)\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right]$$
$$= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N Re\left[\left(W_{X,i}^* - W_{V,i}^*\right)\mathbf{M}_{W_{F^0}}W_{\varepsilon,i}\right] + o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0\right)\right) + o_p(1).$$

Proof of Theorem 4.2. Let $\tilde{\delta} \equiv \tilde{\beta} - \beta^0$. Recall that

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

where $W_{Y,i} = W_{X,i}\beta^0 + \tilde{W}_{F^0}\tilde{\lambda}_i + W_{\varepsilon,i}$. Then

$$\left[\frac{1}{NT}\sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{X,i}\right)\right]\left(\tilde{\beta}-\beta^0\right)$$
$$= \frac{1}{NT}\sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\tilde{W}_{F^0}\tilde{\lambda}_i\right) + \frac{1}{NT}\sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right). \tag{A.5}$$

First, we study $\frac{1}{NT}\sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{F^0}\lambda_i\right)$. 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 that 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

$$\hat{W}_F V_{NL} = \frac{1}{NT}\sum_{i=1}^N W_{X,i}\tilde{\delta}\tilde{\delta}'W_{X,i}^*\hat{W}_F - \frac{1}{NT}\sum_{i=1}^N W_{X,i}\tilde{\delta}\lambda_i'W_{F^0}^*\hat{W}_F - \frac{1}{NT}\sum_{i=1}^N W_{X,i}\tilde{\delta}W_{\varepsilon,i}^*\hat{W}_F^*$$
$$- \frac{1}{NT}\sum_{i=1}^N W_{F^0}\lambda_i\tilde{\delta}'W_{X,i}^*\hat{W}_F - \frac{1}{NT}\sum_{i=1}^N W_{\varepsilon,i}\tilde{\delta}'W_{X,i}^*\hat{W}_F + \frac{1}{NT}\sum_{i=1}^N W_{F^0}\lambda_i W_{\varepsilon,i}^*\hat{W}_F$$

$$+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}\lambda_i'W_{F^0}^*\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}W_{\varepsilon,i}^*\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F^0}\tilde{\lambda}_i\tilde{\lambda}_i'\tilde{W}_{F^0}^*\hat{W}_F$$

$$\equiv\tilde{I}_1+\cdots+\tilde{I}_9. \tag{A.6}$$

This, in conjunction with the definition of $\tilde{H}$ given in Lemma A.6, implies that

$$\tilde{W}_{F^0}-\hat{W}_F\tilde{H}^{-1}=-\left(\tilde{I}_1+\cdots+\tilde{I}_8\right)\left(\tilde{W}_{F^0}^*\hat{W}_F/T\right)^{-1}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)^{-1}\equiv-\left(\tilde{I}_1+\cdots+\tilde{I}_8\right)\tilde{G}.$$

Then

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\tilde{W}_{F^0}\tilde{\lambda}_i\right)&=\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\tilde{W}_{F^0}\tilde{\Gamma}_F^{-1}\lambda_i\right)\\&=\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\left(\tilde{W}_{F^0}-\hat{W}_F\tilde{H}^{-1}\right)\tilde{\Gamma}_F^{-1}\lambda_i\right)\\&=-\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\left(\tilde{I}_1+\cdots+\tilde{I}_8\right)\check{G}\lambda_i\right)\\&\equiv\tilde{J}_1+\cdots+\tilde{J}_8,\end{aligned} \tag{A.7}$$

where $\check{G}=\left(\tilde{W}_{F^0}^*\hat{W}_F/T\right)^{-1}\tilde{\Gamma}_F^{-1}\left(\Lambda'\Lambda/N\right)^{-1}$. 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

$$\begin{aligned}\left\|\tilde{J}_1\right\|&=\left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^*\mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}^*\tilde{\delta}\tilde{\delta}'W_{X,k}\hat{W}_F^*\right)\check{G}\lambda_i\right\|\\&\lesssim\left\|\tilde{G}\right\|\left\|\tilde{\delta}\right\|^2\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|W_{X,i}\|\,\|\lambda_i\|\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2\\&\lesssim\gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|^2\left\{\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\right\}^{3/2}\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\\&=\gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|^2O_p\left(\gamma_L^{3/2-3d_{X,\max}}\right)O_p(1)=O_p\left(\gamma_L^{1+d_{F,\min}-3d_{X,\max}}\left\|\tilde{\delta}\right\|^2\right),\end{aligned}$$

where we use the fact that $\left\|\mathbf{M}_{\hat{W}_F}\right\|_{\mathrm{sp}}=1$ and $\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2=O_p\left(\gamma_L^{1-2d_{X,\max}}\right)$ by Assumption B*(i). It is easy to see that we can express $\tilde{J}_1=-\tilde{J}_1^*\tilde{\delta}$ for some matrix $\tilde{J}_1^*$ with $\left\|\tilde{J}_1^*\right\|=O_p\left(\gamma_L^{1+d_{F,\min}-3d_{X,\max}}\left\|\tilde{\delta}\right\|\right)$. As in the time domain, $\tilde{J}_2$ will enter the asymptotic distribution and it is therefore retained for now. For $\tilde{J}_3$, we make the following decomposition:

$$\tilde{J}_3=\frac{-1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}\tilde{\delta}W_{\varepsilon,k}^*\hat{W}_F\right)\check{G}\lambda_i\right)$$

$$= \frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathrm{Re}\left[\left(\frac{W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}W_{X,k}}{T}\right)\left(\frac{W_{\varepsilon,k}^{*}\tilde{W}_{F^0}\tilde{H}}{T}\right)\check{G}\lambda_i\tilde{\delta}\right]$$
$$+\frac{-1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathrm{Re}\left[\left(\frac{W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}W_{X,k}}{T}\right)\left(\frac{W_{\varepsilon,k}^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)}{T}\right)\check{G}\lambda_i\tilde{\delta}\right] \equiv \tilde{J}_{3,1}+\tilde{J}_{3,2}.$$

First,

$$\left\|\tilde{J}_{3,1}\right\| \lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\left[\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{W_{\varepsilon,k}^{*}\tilde{W}_{F^0}\tilde{H}}{T}\right\|^2\right]^{\frac{1}{2}}$$
$$\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\left[\frac{1}{NT^2}\sum_{k=1}^{N}\left\|W_{\varepsilon,k}^{*}\tilde{W}_{F^0}\right\|^2\right]^{\frac{1}{2}}\left\|\tilde{H}\right\|$$
$$= \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|O_p\left(\gamma_L^{1-2d_{X,\max}}\right)O_p(1)O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p(1)O_p\left(\gamma_L^{1-2d_{F,\max}}\right)$$
$$= O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{2-2d_{F,\max}-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\left\|\tilde{\delta}\right\|\right)$$

by Assumption B(iv), B*(i), C*(i) and D*(iii). By the same reasoning and Lemma A.6,

$$\left\|\tilde{J}_{3,2}\right\| \lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\left[\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right]^{\frac{1}{2}}\left[\frac{1}{NT}\sum_{k=1}^{N}\|W_{\varepsilon,k}\|^2\right]^{\frac{1}{2}}\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|$$
$$= O_p\left(\gamma_L^{1-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\left\|\tilde{\delta}\right\|\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right),$$

where $\delta_{W1,NT}$ is defined in Lemma A.6. In addition, it is easy to see that we can write $\tilde{J}_{3,l}=-\tilde{J}_{3,l}^{*}\tilde{\delta}$ for some matrix $\tilde{J}_{3,l}^{*}$ for $l=1,2$ with $\left\|\tilde{J}_{3,1}^{*}\right\| = O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{2-2d_{F,\max}-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\right)$ and $\tilde{J}_{3,2}^{*}=O_p(\gamma_L^{2-2d_{X,\max}-d_{F,\max}+\left(d_{F,\min}-2d_\varepsilon\right)}N^{-\frac{1}{2}}+\gamma_L^{1-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\delta_{W1,NT}\left\|\tilde{\delta}\right\|)$.

Next,

$$\left\|\tilde{J}_4\right\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N}W_{F^0}\lambda_k\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right)\check{G}\lambda_i\right\|$$
$$= \left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^{N}\left(\tilde{W}_{F^0}-\hat{W}_F\tilde{H}^{-1}\right)\tilde{\lambda}_k\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right)\check{G}\lambda_i\right\|$$
$$\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\left(\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\|W_{X,i}\|\,\|\lambda_i\|\right)^2\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F\right\|\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\left\|\tilde{H}^{-1}\Gamma_F^{-1}\right\|$$
$$\lesssim \gamma_L^{2d_{F,\min}-1}\left\|\tilde{\delta}\right\|\frac{1}{NT}\sum_{i=1}^{N}\|W_{X,i}\|^2\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|$$
$$= O_p\left(\gamma_L^{2d_{F,\min}-1}\left\|\tilde{\delta}\right\|\right)O_p\left(\gamma_L^{1-2d_{X,\max}}\right)O_p(1)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right)$$

$$= O_p\left(\gamma_L^{2d_{F,\min}-2d_{X,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\left\|\tilde{\delta}\right\|\right),$$

and

$$\begin{aligned}
\left\|\tilde{J}_5\right\| &= \left\|\frac{1}{NT}\sum_{i=1}^N W_{X,i}^* \mathbf{M}_{\hat{W}_F}\left(\frac{1}{NT}\sum_{k=1}^N W_{\varepsilon,k}\tilde{\delta}' W_{X,k}^* \hat{W}_F\right)\breve{G}\lambda_i\right\| \\
&\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\| \frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^N \|W_{X,i}\|\|\lambda_i\| \frac{1}{NT}\sum_{k=1}^N \|W_{\varepsilon,k}\|\|W_{X,k}\| \frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F\right\| \\
&= O_p\left(\gamma_L^{d_{F,\min}-\frac{1}{2}}\left\|\tilde{\delta}\right\|\right) O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right) O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right) = O_p\left(\gamma_L^{1-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\left\|\tilde{\delta}\right\|\right).
\end{aligned}$$

It is easy to see that we can $\tilde{J}_l = -\tilde{J}_l^*\tilde{\delta}$ for $l = 4, 5$ with $\left\|\tilde{J}_4^*\right\| = O_p(\gamma_L^{2d_{F,\min}-2d_{X,\max}} N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}$ $+\gamma_L^{2d_{F,\min}-2d_{X,\max}}\delta_{W1,NT}\left\|\tilde{\delta}\right\|)$ and $\left\|\tilde{J}_5^*\right\| = O_p(\gamma_L^{1-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)})$. For $\tilde{J}_6$, we have by Lemma A.6 and A.7 that

$$\begin{aligned}
\left\|\tilde{J}_6\right\| &= \left\|\frac{1}{NT}\sum_{i=1}^N W_{X,i}^* \mathbf{M}_{\hat{W}_F}\tilde{W}_{F^0}\tilde{\lambda}_k\left(\frac{1}{N}\sum_{k=1}^N \frac{W_{\varepsilon,k}^*\hat{W}_F}{T}\breve{G}\lambda_i\right)\right\| \\
&= \left\|\frac{1}{NT}\sum_{i=1}^N W_{X,i}^* \mathbf{M}_{\hat{W}_F}\left(\tilde{W}_{F^0} - \hat{W}_F\tilde{H}^{-1}\right)\tilde{\Gamma}_F^{-1}\left(\frac{1}{N}\sum_{k=1}^N \lambda_k\frac{W_{\varepsilon,k}^*\hat{W}_F}{T}\right)\breve{G}\lambda_i\right\| \\
&\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^N \|W_{X,i}\|\|\lambda_i\| \frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\| \left\|\frac{1}{N}\sum_{k=1}^N \lambda_k\frac{W_{\varepsilon,k}^*\hat{W}_F}{T}\right\| \left\|\tilde{H}^{-1}\tilde{\Gamma}_F^{-1}\right\| \\
&= O_p\left(\gamma_L^{2d_{F,\min}-1}\right) O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right) O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right) \\
&\times O_p\left(\delta_{W,NL}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right) \\
&= O_p\left(\Delta_{1,NT}\left\|\tilde{\delta}\right\|^2 + \Delta_{2,NT}\left\|\tilde{\delta}\right\| + \Delta_{3,NT}\right),
\end{aligned}$$

where $\Delta_{1,NT} = \gamma_L^{\frac{1}{2}-d_{X,\max}+2d_{F,\min}-1}\delta_{W1,NT}\delta_{W,NL}$, $\Delta_{2,NT} = \gamma_L^{\frac{1}{2}-d_{X,\max}+2d_{F,\min}-1}(\delta_{W1,NT}N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}$ $+\delta_{W,NL}N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon})$, and $\Delta_{3,NT} = N^{-1}L^{-\frac{1}{2}}\gamma_L^{2-d_{X,\max}+2d_{F,\min}-3d_{F,\max}-2d_\varepsilon}$. Note that we can write $\tilde{J}_6 = -\tilde{J}_{6,1}^*\tilde{\delta} + \tilde{J}_{6,2}^*$ with $\left\|\tilde{J}_{6,1}^*\right\| = O_p\left(\Delta_{1,NT}\left\|\tilde{\delta}\right\| + \Delta_{2,NT}\right)$ and $\left\|\tilde{J}_{6,2}^*\right\| = O_p\left(\Delta_{3,NT}\right)$. Next, $\tilde{J}_7$ contributes to the asymptotic distribution and is kept here.

Last, for $\tilde{J}_8$, by Lemma A.8 we have $\tilde{J}_8 = A_{NT} + \breve{J}_8$, where

$$\begin{aligned}
A_{NT} &= -\frac{1}{NT}\sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^N diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\breve{G}\lambda_i\right) \\
&= O_p\left(\frac{1}{L}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right),
\end{aligned} \tag{A.8}$$

and

$$\left\|\breve{J}_8\right\| = O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$$

$$+ O_p\left(\left(T^{2d_\varepsilon-1}\gamma_L^{d_{F,\min}-d_{X,\max}} + N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon+\left(d_{F,\min}-d_{X,\max}\right)}\right)\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)$$
$$\equiv O_p\left(\check{\Delta}_{1,NT}\left\|\tilde{\delta}\right\| + \check{\Delta}_{2,NT}\right).$$

Here A_{NT} will enter the bias and we can write $\check{J}_8 = -\check{J}^*_{8,1}\tilde{\delta} + \check{J}^*_{8,2}$ with $\left\|\check{J}^*_{8,1}\right\| = O_p\left(\check{\Delta}_{1,NT}\right)$ and $\left\|\check{J}^*_{8,2}\right\| = O_p\left(\check{\Delta}_{2,NT}\right)$.

Now, by (A.5) and (A.7), we have

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,i}\right)\right]\left(\tilde{\beta}-\beta^0\right) = \left(\tilde{J}_1+\cdots+\tilde{J}_8\right) + \frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right).$$

It follows that

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,i}\right) - \tilde{J}_2 + \tilde{J}_*\right]\left(\tilde{\beta}-\beta^0\right)$$
$$= \tilde{J}^*_{6,2} + A_{NT} + \check{J}_{8,2} + \frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right) + \tilde{J}_7, \tag{A.9}$$

where $\tilde{J}_* = \tilde{J}^*_1+\tilde{J}^*_{3,1}+\tilde{J}^*_{3,2}+\tilde{J}^*_4+\tilde{J}^*_5+\tilde{J}^*_{6,1}$. By construction, $D^\dagger_{NL}\left(\hat{W}_F\right) = \frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,i}\right)$ $-\tilde{J}_2$. Let $\hat{C}_{NL} = \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}W^*_{X,k}\mathbf{M}_{\hat{W}_F}\right)W_{\varepsilon,i}\right)$. Then premultiplying both sides of (A.9) by $\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z$ yields

$$\gamma_L^{-1}\Gamma_Z\left[D^\dagger_{NL}\left(\hat{W}_F\right)+\tilde{J}_*\right]\Gamma_Z\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0\right) = \sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z\left(\tilde{J}^*_{6,2}+A_{NT}+\check{J}_{8,2}\right)+\hat{C}_{NL}. \tag{A.10}$$

To simplify the last expression and obtain the final distributional result, we will show that: (i) $\gamma_L^{-1}\Gamma_Z\left[D^\dagger_{NL}\left(\hat{W}_F\right)-D^\dagger_{NL}\right]\Gamma_Z = o_p(1)$; (ii) $\left\|\tilde{J}_*\right\| = o_p\left(\gamma_L^{1-2d_{Z,\min}}\right)$; (iii) $\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z\tilde{J}^*_{l,2} = o_p(1)$ for $l=6,8$; and (iv) $\hat{C}_{NL} = C_{NL} + C^*_{NL}\tilde{\delta} + o_p(1)$ where

$$\begin{aligned} C_{NL} &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\left(W^*_{X,i}\mathbf{M}_{W_{F^0}} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}W^*_{X,k}\mathbf{M}_{W_{F^0}}\right)W_{\varepsilon,i}\right) \\ &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W^*_{Z,i}W_{\varepsilon,i}\right), \end{aligned}$$

and $C^*_{NL} = o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z\right\|^{-1}\right)$. Noting (ii) implies that $\gamma_L^{-1}\Gamma_Z\tilde{J}_*\Gamma_Z = o_p(1)$, combining these results with (A.10) yields

$$\left[\gamma_L^{-1}\Gamma_Z D^\dagger_{NL}\Gamma_Z + o_p(1)\right]\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0\right) = C_{NL} + \sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z A_{NT} + o_p(1).$$

Let $D^W_{NL} = \gamma_L^{-1}\Gamma_Z D^\dagger_{NL}\Gamma_Z + o_p(1)$. Then by Assumption D(i), $D^W_{NL} \xrightarrow{p} D^W_0 > 0$. It follows that

$$\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\tilde{\beta}-\beta^0\right) = \left[D^W_{NL}+o_p(1)\right]^{-1}C_{NL} + \left[D^W_{NL}+o_p(1)\right]^{-1}\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z A_{NT} + o_p(1).$$

Then

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

where A_{NT}^W is as defined in the statement of Theorem 4.2 and the second line holds by Assumptions D(i) and D*(ii) and the Slutsky theorem. In addition, we can show that $A_{NT}^W=O_p(\phi_L)$ by (A.8), where $\phi_L=L^{-1}\gamma_L^{2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}$.

We now show (i)-(iv) in turn. For (i), the result holds by Lemma A.6 and our regularity conditions. For (ii), we can apply Assumption E* about the relative magnitude among the memory parameters and the convergence rate of $\tilde{\delta}$ to show that

$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_1^*\right\| = O_p\left(\gamma_L^{2d_{Z,\min}+d_{F,\min}-3d_{X,\max}}\left\|\tilde{\delta}\right\|\right)=o_p(1);$$
$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_{3,1}^*\right\| = O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{1+2d_{Z,\min}-2d_{F,\max}-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\right)=o_p(1);$$
$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_{3,2}^*\right\| = O_p\left(\gamma_L^{2d_{Z,\min}-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)=o_p(1);$$
$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_4^*\right\| = O_p\left(\gamma_L^{2d_{Z,\min}+2d_{F,\min}-2d_{X,\max}-1}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)=o_p(1);$$
$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_5^*\right\| = O_p\left(\gamma_L^{2d_{Z,\min}-2d_{X,\max}+\left(d_{F,\min}-d_\varepsilon\right)}\left\|\tilde{\delta}\right\|\right)=o_p(1);$$
$$\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_{6,1}^*\right\| = \gamma_L^{2d_{Z,\min}-1}O_p\left(\Delta_{1,NT}\left\|\tilde{\delta}\right\|+\Delta_{2,NT}\right)=O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}+2d_{F,\min}-\frac{3}{2}}\delta_{W1,NT}\delta_{W,NL}\left\|\tilde{\delta}\right\|\right)$$
$$+O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}+2d_{F,\min}-\frac{3}{2}}(\delta_{W1,NT}N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}+\delta_{W,NL}N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon})\right)$$
$$= o_p(1);$$

and

$$\gamma_L^{2d_{Z,\min}-1}\left\|\check{J}_{8,1}^*\right\|=O_p\left(\delta_{W1,NT}\left(T^{2d_\varepsilon-1}\gamma_L^{d_{F,\min}-d_{X,\max}+2d_{Z,\min}-1}+N^{-\frac{1}{2}}\gamma_L^{2d_{Z,\min}-2d_\varepsilon+\left(d_{F,\min}-d_{X,\max}\right)}\right)\right)=o_p(1).$$

Then $\gamma_L^{2d_{Z,\min}-1}\left\|\tilde{J}_1^*\right\|=o_p(1)$. Next, we show (iii). Following the above analyses of $\tilde{J}_6$ and $\tilde{J}_8$, we have

$$\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\left\|\tilde{J}_{6,2}^*\right\|=O_p\left(\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\Delta_{3,NT}\right)$$
$$=O_p\left(\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\gamma_L^{\frac{1}{2}-d_{X,\max}+2d_{F,\min}-1}N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)$$
$$+O_p\left(N^{-\frac{1}{2}}\gamma_L^{1+2d_{F,\min}+d_{Z,\min}-3d_{F,\max}-d_{X,\max}-d_\varepsilon}\right)=o_p(1),$$

and

$$\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\left\|\check{J}_{8,2}^*\right\|=O_p\left(\gamma_L^{\frac{3}{2}+d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-d_\varepsilon}\right)$$

$$+ O_p\left(\frac{L^{\frac{1}{2}}}{T^{1-2d_\varepsilon}}\gamma_L^{d_{F,\min}+d_{Z,\min}-d_{X,\max}-d_{F,\max}} + \gamma_L^{\frac{3}{2}-2d_\varepsilon+d_{Z,\min}+d_{F,\min}-d_{X,\max}-d_{F,\max}}\right)$$
$$= o_p(1).$$

Then (iii) holds. Last, (iv) follows from Lemma A.9. This completes the proof of Theorem 4.2. ■

Proof of Theorem 4.3. Recall that $\hat{\Sigma}^W_{NL} = \frac{1}{N^2T^2}\sum_{i=1}^N \mathrm{Re}\left(\hat{W}^*_{Z,i}\hat{W}_{\varepsilon,i}\right)\mathrm{Re}\left(\hat{W}^*_{\varepsilon,i}\hat{W}_{Z,i}\right)$ and $\Sigma^\dagger_{NL} = \frac{\gamma_L^{2d_\varepsilon-1}}{NT}\sum_{i=1}^N \mathrm{Re}\left(\Gamma_Z W^*_{Z,i}W_{\varepsilon,i}\right)\mathrm{Re}\left(W^*_{\varepsilon,i}W_{Z,i}\Gamma_Z\right)$. It suffices to prove the theorem by showing that (i) $\gamma_L^{-1}\Gamma_Z(\hat{D}^W_{NL} - D^\dagger_{NL})\Gamma_Z = o_p(1)$, (ii) $NL\gamma_L^{2d_\varepsilon-2}\Gamma_Z\hat{\Sigma}^W_{NL}\Gamma_Z - \Sigma^\dagger_{NL} = o_p(1)$, and (iii) $\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}(\hat{A}^W_{NT} - A^W_{NT}) = o_p(1)$.

We first prove (i). Recall that $\hat{D}^W_{NL} = \frac{1}{NT}\sum_{i=1}^N \mathrm{Re}\left(\hat{W}^*_{Z,i}\hat{W}_{Z,i}\right) = \frac{1}{NT}\sum_{i=1}^N \mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,i}\right) - \frac{1}{N^2T}\sum_{i=1}^N\sum_{k=1}^N \mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,k}\widehat{a}_{ik}\right)$ and $D^\dagger_{NL}$ has equal to $\hat{D}^W_{NL}$ with $\hat{W}_F$ and $\widehat{a}_{ik}$ replaced by their true values. It follows that

$$\begin{aligned}
\gamma_L^{-1}\Gamma_Z\left(\hat{D}^W_{NL} - D^\dagger_{NL}\right)\Gamma_Z &= \frac{\gamma_L^{-1}\Gamma_Z}{NT}\sum_{i=1}^N \mathrm{Re}\left(W^*_{X,i}\left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{\tilde{W}_{F^0}}\right)W_{X,i}\right)\Gamma_Z \\
&\quad - \frac{\gamma_L^{-1}\Gamma_Z}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \mathrm{Re}\left(W^*_{X,i}\mathbf{M}_{\hat{W}_F}W_{X,k}\left(\widehat{a}_{ik} - a_{ik}\right)\right)\right]\Gamma_Z \\
&\quad - \frac{\gamma_L^{-1}\Gamma_Z}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \mathrm{Re}\left(W^*_{X,i}\left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{\tilde{W}_{F^0}}\right)W^*_{X,k}a_{ik}\right)\right]\Gamma_Z \\
&\equiv d_1 + d_2 + d_3.
\end{aligned}$$

where we use the fact that $\mathbf{M}_{\tilde{W}_{F^0}} = \mathbf{M}_{W_{F^0}}$. For d_1, we have

$$\|d_1\| \lesssim \gamma_L^{2d_{Z,\min}-1}\frac{1}{NT}\sum_{i=1}^N \|W_{X,i}\|^2\left\|\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right\| \lesssim \gamma_L^{2d_{Z,\min}-1}\left\|\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right\|.$$

Note that

$$\begin{aligned}
&\left\|\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right\|^2/2 \\
= \;& \mathrm{tr}\left(I_R - \hat{W}^*_F\mathbf{P}_{\tilde{W}_{F^0}}\hat{W}_F/T\right) = 2\mathrm{tr}\left(\mathbb{I}_R - \frac{\hat{W}^*_F\tilde{W}_{F^0}}{T}\left(\frac{\tilde{W}^*_{F^0}\tilde{W}_{F^0}}{T}\right)^{-1}\frac{\tilde{W}^*_{F^0}\hat{W}_F}{T}\right) \\
= \;& \mathrm{tr}\left(\mathbb{I}_R - \left[\frac{\tilde{W}^*_{F^0}\tilde{W}_{F^0}\tilde{H}}{T} + \frac{\tilde{W}^*_{F^0}(\hat{W}_F - \tilde{W}_{F^0}\tilde{H})}{T}\right]^*\left(\frac{\tilde{W}^*_{F^0}\tilde{W}_{F^0}}{T}\right)^{-1}\left[\frac{\tilde{W}^*_{F^0}\tilde{W}_{F^0}\tilde{H}}{T} + \frac{\tilde{W}^*_{F^0}(\hat{W}_F - \tilde{W}_{F^0}\tilde{H})}{T}\right]\right) \\
= \;& \mathrm{tr}\Bigg\{\left[\mathbb{I}_R - \tilde{H}^*\left(\frac{\tilde{W}^*_{F^0}\tilde{W}_{F^0}}{T}\right)\tilde{H}\right] - \frac{1}{T}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)^*\tilde{W}_{F^0}\tilde{H} - \frac{1}{T}\tilde{H}^*\tilde{W}^*_{F^0}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right) \\
&-\frac{1}{T^2}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)^*\tilde{W}_{F^0}\left(\frac{\tilde{W}^*_{F^0}\tilde{W}_{F^0}}{T}\right)^{-1}\tilde{W}^*_{F^0}\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\Bigg\}
\end{aligned}$$

$$\equiv \quad d_{11} + d_{12} + d_{13} + d_{14}.$$

Note that $\left\|\tilde{H}\right\| = O_p(\gamma_L^{1-2d_{F,\max}})$ as used in the proof of Lemma A.6. For d_{11}, we have

$$\begin{aligned} d_{11} &= \operatorname{tr}\left(\mathbb{I}_R - \tilde{H}^*\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)\tilde{H}\right) \le \sqrt{R}\left\|\mathbb{I}_R - \tilde{H}^*\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)\tilde{H}\right\| \\ &= O_p\left(\gamma_L^{1-2d_{F,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right) = o_p\left(\gamma_L^{1-2d_{Z,\min}}\right), \end{aligned}$$

by (B.10) in the proof of Lemma A.9, and Assumption E*(iii) given the convergence rate of $\tilde{\beta}$. Similarly, by Lemma A.6 and Assumption B*(iii), $|d_{12}| = |d_{13}| = o_p\left(\gamma_L^{1-2d_{Z,\min}}\right)$, and $|d_{14}|$ is of smaller order. Consequently, we have $\|d_1\| \lesssim \gamma_L^{2d_{Z,\min}-1}\left\|\mathbf{P}_{\hat{W}_F} - \mathbf{P}_{\tilde{W}_{F^0}}\right\| = o_p(1)$. By Assumption B(iv), $\|d_3\|$ has the same order as $\|d_1\|$.

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 $\hat{a}_{ik} = \hat{\lambda}_k^*\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i$, where $\hat{\Lambda} = \overline{\hat{W}_U}\hat{W}_F/T$, $\hat{W}_U \equiv (\hat{W}_{U,1}, ..., \hat{W}_{U,N})'$ is an $N \times L$ complex matrix with $\hat{W}_{U,i} = W_{Y,i} - W_{X,i}\tilde{\beta}$, and $\hat{\lambda}_i = \hat{W}_F^*\hat{W}_{U,i}/T$. Note that

$$\begin{aligned} \hat{a}_{ik} - a_{ik} &= \hat{\lambda}_k^*\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right) + \hat{\lambda}_k^*\left[\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1} - \tilde{H}^*\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{H}\right]\tilde{H}^{-1}\tilde{\lambda}_i \\ &\quad + \left(\hat{\lambda}_k - \tilde{H}^{-1}\tilde{\lambda}_k\right)^*\tilde{H}^*\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{\lambda}_i \equiv \sum_{\ell=1}^{3}\vartheta_{\ell,ik}. \end{aligned}$$

Then $d_2 = -\sum_{\ell=1}^3 \frac{\gamma_L^{-1}\Gamma_Z}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \operatorname{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{X,k}\vartheta_{\ell,ik}\right)\right]\Gamma_Z \equiv \sum_{\ell=1}^3 d_{2\ell}$. For d_{21}, we have

$$\begin{aligned} \|d_{21}\| &= \left\|\frac{\gamma_L^{-1}\Gamma_Z}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \operatorname{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{X,k}\hat{\lambda}_k^*\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right)\right)\right]\Gamma_Z\right\| \\ &\lesssim \gamma_L^{2d_{Z,\min}-1}\frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^N\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right)W_{X,i}^*\right\|\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^N\|W_{X,k}\|\left\|\hat{\lambda}_k^*\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\right\| \\ &\lesssim \gamma_L^{2d_{Z,\min}-d_{X,\max}-\frac{1}{2}}\frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^N\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right)W_{X,i}^*\right\|, \end{aligned}$$

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 the frequency domain. Note that

$$\begin{aligned} \hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i &= \hat{W}_F^*\left(W_{Y,i} - W_{X,i}\tilde{\beta}\right)/T - \tilde{H}^{-1}\tilde{\lambda}_i \\ &= \frac{1}{T}\hat{W}_F^*\left(\tilde{W}_{F^0}\tilde{\lambda}_i + W_{\varepsilon,i} - W_{X,i}\tilde{\delta}\right) - \tilde{H}^{-1}\tilde{\lambda}_i \\ &= \frac{1}{T}\hat{W}_F^*\left(\tilde{W}_{F^0} - \hat{W}_F\tilde{H}^{-1}\right)\tilde{\lambda}_i + \frac{1}{T}\hat{W}_{F^0}^*W_{\varepsilon,i} - \frac{1}{T}\hat{W}_F^*W_{X,i}\tilde{\delta} \end{aligned}$$

where we use the fact that $\frac{1}{T}\hat{W}_F^*\hat{W}_F = \mathbb{I}_R$. So

$$\frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right)W_{X,i}^*\right\|$$
$$= O_p\left(\left(\gamma_L^{d_{F,\min}-d_{X,\max}}\delta_{W1,NT} + \gamma_L^{1-2d_{X,\max}}\right)\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1+d_{F,\min}-d_{F,\max}-d_{X,\max}-d_\varepsilon} + N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right) \tag{A.11}$$

by Lemma A.6 and D*(iii) and Assumption E*(iii). Therefore

$$\|d_{21}\| = O_p\left(D_{1,NT}\left\|\tilde{\delta}\right\| + D_{2,NT}\right) = o_p(1),$$

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 $D_{2,NT} = N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}+2d_{Z,\min}-2d_{X,\max}-d_\varepsilon}$ $+N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}+d_{F,\min}+2d_{Z,\min}-2d_{X,\max}-d_{F,\max}-d_\varepsilon}$, where the last equality in the above displayed equation holds by the convergence rate of $\tilde{\beta}$ and Assumption E*(iii). Analogously, we can show that $\|d_{21}\| = o_p(1)$. For d_{22}, we have

$$\begin{aligned}\|d_{22}\| &= \left\|\frac{\gamma_L^{-1}\Gamma_Z}{T}\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\text{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{X,k}\hat{\lambda}_k^*\tilde{d}_{22}\tilde{H}^{-1}\tilde{\lambda}_i\right)\right]\Gamma_Z\right\| \\ &\lesssim \gamma_L^{2d_{Z,\min}-1}\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\|W_{X,i}\|\left\|\tilde{H}^{-1}\tilde{\lambda}_i\right\|\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\|W_{X,k}\|\left\|\hat{\lambda}_k\right\|\left\|\tilde{d}_{22}\right\| \\ &= O_p\left(\gamma_L^{2d_{F,\min}+2d_{Z,\min}-2d_{X,\max}-1}\right)\left\|\tilde{d}_{22}\right\|,\end{aligned}$$

where $\tilde{d}_{22} = \left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1} - \tilde{H}^*\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{H}$. Analogous to the analysis of (A.11), we have

$$\begin{aligned}\left\|\tilde{d}_{22}\right\| &= \left\|\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\left(\tilde{H}^{-1}\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\tilde{H}^{*-1} - \frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)\tilde{H}^*\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{H}\right\| \\ &= \left\|\left(\frac{\hat{\Lambda}^*\hat{\Lambda}}{N}\right)^{-1}\left[\frac{1}{N}\left(\hat{\Lambda}^* - \tilde{H}^{-1}\tilde{\Lambda}^*\right)\tilde{\Lambda}\tilde{H}^{*-1} + \frac{1}{N}\hat{\Lambda}^*\left(\hat{\Lambda} - \tilde{\Lambda}\tilde{H}^{*-1}\right)\right]\tilde{H}^*\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\tilde{H}\right\| \\ &\lesssim \gamma_L^{\frac{5}{2}+d_{F,\min}-6d_{F,\max}}\frac{1}{\sqrt{N}}\left\|\hat{\Lambda}^* - \tilde{H}^{-1}\tilde{\Lambda}^*\right\| \\ &\leq \gamma_L^{\frac{5}{2}+d_{F,\min}-6d_{F,\max}} \\ &\times O_p\left(\left(\gamma_L^{d_{F,\min}-\frac{1}{2}}\delta_{W1,NT} + \gamma_L^{\frac{1}{2}-d_{X,\max}}\right)\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}+d_{F,\min}-d_{F,\max}-d_\varepsilon} + L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right).\end{aligned} \tag{A.12}$$

Therefore d_{22} is dominated by d_{21} and thus asymptotically negligible as well.

Next, we show (ii). Noting that $NL\gamma_L^{-1} = NT/(2\pi)$, we have

$$NL\gamma_L^{2d_\varepsilon-2}\Gamma_Z\hat{\Sigma}_{NL}^W\Gamma_Z - \Sigma_{NL}^\dagger$$

$$= \frac{\gamma_L^{2d_\varepsilon - 1}}{2\pi NT} \sum_{i=1}^{N} \Gamma_Z \left[\text{Re}\left(\hat{W}_{Z,i}^* \hat{W}_{\varepsilon,i} \right) \text{Re}\left(\hat{W}_{\varepsilon,i}^* \hat{W}_{Z,i} \right) - \text{Re}\left(W_{Z,i}^* W_{\varepsilon,i} \right) \text{Re}\left(W_{\varepsilon,i}^* W_{Z,i} \right) \right] \Gamma_Z$$

$$= \frac{\gamma_L^{2d_\varepsilon - 1}}{2\pi NT} \sum_{i=1}^{N} \Gamma_Z \text{Re}\left(\hat{W}_{Z,i}^* \hat{W}_{\varepsilon,i} - W_{Z,i}^* W_{\varepsilon,i} \right) \text{Re}\left(W_{\varepsilon,i}^* W_{Z,i} \right) \Gamma_Z$$

$$+ \frac{\gamma_L^{2d_\varepsilon - 1}}{2\pi NT} \sum_{i=1}^{N} \Gamma_Z \text{Re}\left(\hat{W}_{Z,i}^* \hat{W}_{\varepsilon,i} \right) \text{Re}\left(\hat{W}_{\varepsilon,i}^* \hat{W}_{Z,i} - W_{\varepsilon,i}^* W_{Z,i} \right) \Gamma_Z \equiv w_1 + w_2.$$

For w_1, we have

$$\|w_1\| \lesssim \left(\frac{\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z \left(\hat{W}_{Z,i}^* \hat{W}_{\varepsilon,i} - W_{Z,i}^* W_{\varepsilon,i} \right) \right\|^2 \right)^{\frac{1}{2}} \left(\frac{\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z W_{Z,i}^* W_{\varepsilon,i} \right\|^2 \right)^{\frac{1}{2}} \equiv \sqrt{w_{11}} \sqrt{w_{12}}.$$

By Assumption D*(iii), $w_{12} = O_p(1)$. For w_{11}, we have

$$w_{11} \leq \frac{2\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z \hat{W}_{Z,i}^* \left(\hat{W}_{\varepsilon,i} - W_{\varepsilon,i} \right) \right\|^2 + \frac{2\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z \left(\hat{W}_{Z,i}^* - W_{Z,i}^* \right) W_{\varepsilon,i} \right\|^2 \equiv w_{11,1} + w_{11,2}.$$

Following the proof in (i), we can show that $w_{11,2} = o_p(1)$. To study $w_{11,1}$, we note that

$$\begin{aligned} \hat{W}_{\varepsilon,i} - W_{\varepsilon,i} &= (W_{Y,i} - W_{\varepsilon,i}) - W_{X,i}\tilde{\beta} - \hat{W}_F \hat{\lambda}_i \\ &= -W_{X,i}\tilde{\delta} + \tilde{W}_{F^0}\tilde{H}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i \right) + \left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H} \right) \tilde{H}^{-1}\tilde{\lambda}_i + \left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H} \right) \left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i \right) \\ &\equiv \sum_{\ell=1}^{4} \tilde{w}_{\ell i}. \end{aligned} \tag{A.13}$$

It follows that

$$\begin{aligned} w_{11,1} &\lesssim \frac{\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z \hat{W}_{Z,i}^* W_{X,i}\tilde{\delta} \right\|^2 + \frac{\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z \hat{W}_{Z,i}^* \tilde{W}_{F^0}\tilde{H}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i \right) \right\|^2 \\ &\quad + \frac{\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z \hat{W}_{Z,i}^* \left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H} \right) \tilde{H}^{-1}\tilde{\lambda}_i \right\|^2 + \frac{\gamma_L^{2d_\varepsilon - 1}}{NT} \sum_{i=1}^{N} \left\| \Gamma_Z \hat{W}_{Z,i}^* \tilde{w}_{4i} \right\|^2 \\ &= o_p\left(\gamma_L^{1-2d_{X,\max}} \right) + o_p\left(\gamma_L^{1-2d_{F,\max}} \right) + O_p\left(\gamma_L^{2d_\varepsilon + 2d_{F,\max} - 1} \left\| \hat{W}_F - \tilde{W}_{F^0}\tilde{H} \right\|^2 \right) \\ &= o_p(1), \end{aligned}$$

by Assumption E*(iii) and (A.11). Then $w_1 = o_p(1)$. An analogous proof gives $w_2 = o_p(1)$.

Now, we show (iii): $\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left(\hat{A}_{NT}^W - A_{NT}^W \right) = o_p(1)$. Recall that

$$\begin{aligned} &\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\hat{A}_{NT}^W \\ &= -\left[\gamma_L^{-1}\Gamma_Z \hat{D}_{NL}^W \Gamma_Z \right]^{-1} \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \text{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \frac{1}{NT} \sum_{k=1}^{N} \hat{\Omega}_k \hat{W}_F \left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N} \right)^{-1} \hat{\lambda}_i \right) \end{aligned}$$

$$\equiv -\left[\gamma_L^{-1}\Gamma_Z \hat{D}_{NL}^W \Gamma_Z\right]^{-1} \mathbb{N}_1,$$

and

$$\begin{aligned}
&\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1} A_{NT}^W \\
&= -\left[\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z\right]^{-1} \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \frac{1}{NT}\sum_{k=1}^N diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F \tilde{G}\tilde{\lambda}_i\right) \\
&= -\left[\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z\right]^{-1} \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \frac{1}{NT}\sum_{k=1}^N diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F V_{NL}^{-1}\tilde{H}^{-1}\tilde{\lambda}_i\right) \\
&\equiv -\left[\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z\right]^{-1} \mathbb{N}_2, ,
\end{aligned}$$

where we use the fact that $\tilde{G}\tilde{H} = V_{NL}^{-1}$. Note that the two denominator parts are $\gamma_L^{-1}\Gamma_Z \hat{D}_{NL}^W \Gamma_Z$ and $\gamma_L^{-1}\Gamma_Z D_{NL}^{\dagger}\Gamma_Z$, and we have shown that their difference is asymptotic negligible above. So it suffices to consider the difference $\mathbb{N}_1$ and $\mathbb{N}_2$. Note that

$$\begin{aligned}
\mathbb{N}_1 - \mathbb{N}_2 &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^N \mathrm{Re}\left[W_{X,i}^* \mathbf{M}_{\hat{W}_F} \frac{1}{NT}\sum_{k=1}^N \left(\hat{\Omega}_k - diag\left(|W_{\varepsilon,kj}|^2\right)\right)\hat{W}_F V_{NL}^{-1}\hat{\lambda}_i\right] \\
&\quad + \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^N \mathrm{Re}\left(W_{X,i}^* \mathbf{M}_{\hat{W}_F} \frac{1}{NT}\sum_{k=1}^N diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F V_{NL}^{-1}\left(\hat{\lambda}_i - \tilde{H}^{-1}\tilde{\lambda}_i\right)\right) \\
&\equiv D_1 + D_2,
\end{aligned}$$

where we use the fact that $\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1} = V_{NL}^{-1}$ by (4.7) and (4.10). For D_1, we have

$$\|D_1\| \lesssim \sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon - d_{X,\max} - \frac{1}{2}} \left\| \frac{1}{NT}\sum_{k=1}^N \left(\hat{\Omega}_k - diag\left(|W_{\varepsilon,kj}|^2\right)\right)\right\|.$$

Since by (A.13),

$$\left(\left|\hat{W}_{\varepsilon,kj}\right|^2 - |W_{\varepsilon,kj}|^2\right) \lesssim \left\|W_{X,kj}'\tilde{\delta}\right\|^2 + \left\|\tilde{W}_{F^0,j}'\tilde{H}\left(\hat{\lambda}_k - \tilde{H}^{-1}\tilde{\lambda}_k\right)\right\|^2 + \left\|\left(\hat{W}_{F,j}' - \tilde{W}_{F^0,j}'\tilde{H}\right)\tilde{H}^{-1}\tilde{\lambda}_k\right\|^2,$$

we have

$$\begin{aligned}
\|D_1\| &\lesssim \sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon - d_{X,\max} - \frac{1}{2}} \frac{1}{NT}\sum_{k=1}^N \sum_{j=1}^L \left|\left|\hat{W}_{\varepsilon,kj}\right|^2 - |W_{\varepsilon,kj}|^2\right| \\
&\le \sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon - d_{X,\max} - \frac{1}{2}} \\
&\times \frac{1}{NT}\sum_{k=1}^N \left(\left\|W_{X,k}\tilde{\delta}\right\|^2 + \left\|\tilde{W}_{F^0}\tilde{H}\left(\hat{\lambda}_k - \tilde{H}^{-1}\tilde{\lambda}_k\right)\right\|^2 + \left\|\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^{-1}\tilde{\lambda}_k\right\|^2\right) \\
&= O_p\left(\sqrt{NL}\gamma_L^{\frac{1}{2}+d_{Z,\min}+d_\varepsilon - 3d_{X,\max}}\left\|\tilde{\delta}\right\|^2\right) + O_p\left(\sqrt{NL}\gamma_L^{\frac{1}{2}+d_{Z,\min}+d_\varepsilon - d_{X,\max} - d_{F,\max}}\right).
\end{aligned}$$

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 asymptotic negligibility of both D_1 and D_2 by the same reasoning as before and Assumption C*(i) and E*(ii). This completes the proof of Theorem 4.3. ■

A.3 Proofs of the results in Section 5

Proof of Theorem 5.1. First we prove the consistency of $\tilde{\beta}_{(R_{\max})}$, extending the proof of Theorem 4.1 in Moon and Weidner (2015) to the frequency domain framework. Note that $\tilde{\beta}_{(R_{\max})} = \arg\min_{\beta\in\mathbb{R}^P} \mathcal{L}_{NT}^{R_{\max}}(\beta)$. In view of the objective function (4.3), and by considering (4.6), (4.7) and (4.8) together with the property of PCA, we have

$$\begin{aligned}
\mathcal{L}_{NT}^{R_{\max}}(\beta) &= \min_{\tilde{\Lambda}\in\mathbb{C}^{N\times R_{\max}},\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}}} \frac{1}{NT}\left\|W_Y - W_X\beta - \tilde{\Lambda}\tilde{W}_F'\right\|^2 \\
&= \min_{\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}}} \frac{1}{NT}\mathrm{tr}\left[(W_Y - W_X\beta)\,\mathbf{M}_{\tilde{W}_F}(W_Y - W_X\beta)^*\right] \\
&= \frac{1}{NT}\sum_{r=R_{\max}+1}^{L} \mu_r\left[(W_Y - W_X\beta)^*(W_Y - W_X\beta)\right]
\end{aligned}$$

subject to the identification restrictions, where $\mu_r(\cdot)$ represents the r-th largest eigenvalue. Note that $W_Y = W_X\beta^0 + \tilde{\Lambda}^0\tilde{W}_{F^0}' + W_\varepsilon$, where we put the superscripts to emphasize the true values. Then

$$\begin{aligned}
\mathcal{L}_{NT}^{R_{\max}}(\beta) &= \min_{\tilde{\Lambda}\in\mathbb{C}^{N\times R_{\max}},\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}}} \frac{1}{NT}\left\|W_X\delta + W_\varepsilon + \tilde{\Lambda}^0\tilde{W}_{F^0}' - \tilde{\Lambda}\tilde{W}_F'\right\|^2 \\
&\geq \min_{\tilde{\Lambda}\in\mathbb{C}^{N\times R_{\max}+R^0},\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}+R^0}} \frac{1}{NT}\left\|W_X\delta + W_\varepsilon - \tilde{\Lambda}\tilde{W}_F'\right\|^2 \\
&= \min_{\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}+R^0}} \frac{1}{NT}\mathrm{tr}\left[(W_X\delta + W_\varepsilon)\,\mathbf{M}_{\tilde{W}_F}(W_X\delta + W_\varepsilon)^*\right] \\
&= \min_{\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}+R^0}} \frac{1}{NT}\left\{\mathrm{tr}\left[(W_X\delta)\,\mathbf{M}_{\tilde{W}_F}(W_X\delta)^*\right] - \mathrm{tr}\left(W_\varepsilon\mathbf{P}_{\tilde{W}_F}W_\varepsilon^*\right) - 2\mathrm{tr}\left[(W_X\delta)\,\mathbf{P}_{\tilde{W}_F}W_\varepsilon^*\right]\right. \\
&\left.+\mathrm{tr}\left(W_\varepsilon W_\varepsilon^*\right) + 2\mathrm{tr}\left[(W_X\delta)\,W_\varepsilon^*\right]\right\} \\
&\geq \frac{1}{NT}\sum_{r=R_{\max}+R^0+1}^{L} \mu_r\left[(W_X\delta)^*(W_X\delta)\right] + \frac{1}{NT}\mathrm{tr}\left(W_\varepsilon W_\varepsilon^*\right) + \frac{1}{NT}2\mathrm{tr}\left[(W_X\delta)\,W_\varepsilon^*\right] \\
&- \frac{1}{NT}\left(2\left(R_{\max}+R^0\right)\|W_\varepsilon\|^2 + 2\left(R_{\max}+R^0\right)\|W_\varepsilon\|\,\|W_X\delta\|\right) \\
&\geq b\,\|\delta\|^2 + \frac{1}{NT}\mathrm{tr}\left(W_\varepsilon W_\varepsilon^*\right) + O_p\left(\gamma_L^{1-2d_\varepsilon}\right) + O_p\left(\|\delta\|\,\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right) \qquad \text{(A.14)}
\end{aligned}$$

by Assumption B*(i), C*(i) and J, where $\delta = \beta - \beta^0$. Next,

$$\mathcal{L}_{NT}^{R_{\max}}\left(\tilde{\beta}_{(R_{\max})}\right) \leq \mathcal{L}_{NT}^{R_{\max}}\left(\beta^0\right) = \min_{\tilde{\Lambda}\in\mathbb{C}^{N\times R_{\max}},\tilde{W}_F\in\mathbb{C}^{L\times R_{\max}}} \frac{1}{NT}\left\|W_\varepsilon + \tilde{\Lambda}^0\tilde{W}_{F^0}' - \tilde{\Lambda}\tilde{W}_F'\right\|^2$$

$$\leq \frac{1}{NT}\left\|W_{\varepsilon}\right\|^{2}=\frac{1}{NT}\operatorname{tr}\left(W_{\varepsilon}W_{\varepsilon}^{*}\right). \tag{A.15}$$

Let $\tilde{\delta}_{(R_{\max})}=\tilde{\beta}_{(R_{\max})}-\beta^{0}$. Then combining (A.14) and (A.15) we have

$$b\left\|\tilde{\delta}_{(R_{\max})}\right\|^{2}+O_{p}\left(\gamma_{L}^{1-2d_{\varepsilon}}\right)+O_{p}\left(\left\|\tilde{\delta}_{(R_{\max})}\right\|\gamma_{L}^{1-d_{X,\max}-d_{\varepsilon}}\right)\leq 0,$$

which implies $\left\|\tilde{\delta}_{(R_{\max})}\right\|=O_{p}\left(\gamma_{L}^{\frac{1}{2}-d_{X,\max}}\right)=o_{p}(1)$ by Assumption E*(ii).

Next, we prove of the consistency of ER estimator. Let $\mu_{NT,j}$ denote the j-th largest eigenvalue of $\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{F^{0\prime}F^{0}}{T}\right)$ for $j\geq 1$. Let $\bar{c}=c_{1}^{2}\left(1+\sqrt{y}\right)^{2}$, $\underline{c}=c_{2}^{2}y^{**}\left(1-\sqrt{by^{*}}\right)^{2}$ and $y^{**}=\lim_{m\to\infty}\frac{N}{M}$. It suffices to show that

(i) $\frac{\tilde{\mu}_{NT,j}}{\tilde{\mu}_{NT,j+1}}=\frac{\mu_{NT,j}}{\mu_{NT,j+1}}+o_{p}(1)=O_{p}(1)$ for $j=1,\ldots,R^{0}-1$;

(ii) $\frac{\tilde{\mu}_{NT,R^{0}}}{\tilde{\mu}_{NT,R^{0}+1}}\geq\frac{\mu_{NT,R^{0}}+O_{p}\left(N^{-\frac{1}{2}}+m^{-1}+\left(\tilde{\beta}_{(R_{\max})}-\beta\right)\right)}{[\bar{c}+o_{p}(1)]/m}\xrightarrow{p}\infty$; and

(iii) $\frac{\tilde{\mu}_{NT,R^{0}+j}}{\tilde{\mu}_{NT,R^{0}+j+1}}\leq\frac{\bar{c}+o_{p}(1)}{\underline{c}+o_{p}(1)}$ for $j=1,\ldots,\lfloor d^{c}m\rfloor-2R^{0}-1$.

As shown in Ahn and Horenstein (2013), all the reasoning in the proof of their Theorem 1 also holds here except that $\tilde{\mu}_{NT,j}$ denotes the j-th largest eigenvalue of $\frac{\tilde{U}\tilde{U}'}{NT}$, where $\tilde{U}=-X\tilde{\delta}_{(R_{\max})}+U=-\sum_{p=1}^{P}X_{p}\tilde{\delta}_{(R_{\max}),p}+U$ by (5.1). Here $\tilde{\delta}_{(R_{\max}),p}$ denotes the pth element of $\tilde{\delta}_{(R_{\max})}$. Following the proof of Lemma A.11 and A.9 in AH, we can see it is sufficient to show that for any $j=1,\ldots,\lfloor d^{c}m\rfloor-R^{0}$,

$$\psi_{j}\left(\frac{\tilde{U}\tilde{U}'}{NT}\right)=\psi_{j}\left(\frac{UU'}{NT}\right)+o_{p}(1).$$

For notational simplicity, we focus on the case when the regressor is a scalar. Note that

$$\begin{aligned}\frac{\tilde{U}\tilde{U}'}{NT}&=\frac{UU'}{NT}+\left[\frac{1}{NT}\left(X\tilde{\delta}_{R_{\max}}\right)\left(X\tilde{\delta}_{R_{\max}}\right)'-\frac{1}{NT}\left(X\tilde{\delta}_{R_{\max}}\right)U'-\frac{1}{NT}U\left(X\tilde{\delta}_{R_{\max}}\right)'\right]\\&\equiv\frac{UU'}{NT}+R_{NT}.\end{aligned}$$

It is easy to show that $\|R_{NT}\|=O_{p}\left(\tilde{\delta}_{R_{\max}}\right)$ under Assumption B and C(i). Then by Weyl's inequality (or Lemma A.5 in AH) we have

$$\psi_{j}\left(\frac{\tilde{U}\tilde{U}'}{NT}\right)\leq\psi_{j}\left(\frac{UU'}{NT}\right)+\psi_{1}\left(R_{NT}\right)=\psi_{j}\left(\frac{UU'}{NT}\right)+O_{p}\left(\tilde{\delta}_{R_{\max}}\right) \tag{A.16}$$

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 $k=1,\ldots,R^{0}$,

$$\sum_{j=1}^{k}\psi_{j}\left(\frac{\tilde{U}\tilde{U}'}{NT}\right)\geq\operatorname{tr}\left(\frac{1}{NT^{2}}\Xi^{k\prime}UU'\Xi^{k}+\frac{1}{T}\Xi^{k\prime}R\Xi^{k}\right)=\sum_{j=1}^{k}\psi_{j}\left(\frac{UU'}{NT}\right)+O_{p}\left(\tilde{\delta}_{R_{\max}}\right), \tag{A.17}$$

as $\left|\operatorname{tr}\left(\frac{1}{T}\Xi^{k\prime}R\Xi^{k}\right)\right|\leq\left\|\frac{1}{\sqrt{T}}\Xi^{k}\right\|\|R\|$. Note that (A.16) and (A.17) hold for arbitrary $j,k=1,\ldots,\lfloor d^{c}m\rfloor-R^{0}$, which implies that $\psi_{j}\left(\frac{\tilde{U}\tilde{U}'}{NT}\right)=\psi_{j}\left(\frac{UU'}{NT}\right)+O_{p}\left(\tilde{\delta}_{R_{\max}}\right)$. The remaining proof will be the same as in AH using $\frac{UU'}{NT}$ and the mock eigenvalue $\tilde{\mu}_{NT,0}$ so that the consistency of our ER estimator is ensured by the consistency of $\tilde{\beta}_{(R_{\max})}$. ■

B Proofs of the Technical Lemmas

Proof of Lemma A.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 (A.2), the fact that $I_9=F^0\frac{\Lambda'\Lambda}{N}\frac{F^{0\prime}\hat{F}}{T}$, and the definitions of H and G, we have

$$\hat{F}V_{NT}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}-F^0=\hat{F}H^{-1}-F^0$$
$$=(I_1+\cdots+I_8)\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}=(I_1+\cdots+I_8)\,G.$$

Then $T^{-\frac{1}{2}}\left\|\hat{F}V_{NT}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}-F^0\right\|=T^{-\frac{1}{2}}\left\|\hat{F}H^{-1}-F^0\right\|\leq T^{-\frac{1}{2}}\left(\|I_1\|+\cdots+\|I_8\|\right)\|G\|.$ 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,

$$T^{-\frac{1}{2}}\|I_1\|=T^{-\frac{1}{2}}\left\|\frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\hat{\delta}'X_i'\hat{F}\right\|\leq\frac{1}{N}\sum_{i=1}^{N}\frac{\|X_i\|^2}{T}\left\|\hat{\delta}\right\|^2T^{-\frac{1}{2}}\left\|\hat{F}\right\|=O_p\left(\left\|\hat{\delta}\right\|^2\right),$$

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,

$$T^{-\frac{1}{2}}\|I_2\|=T^{-\frac{1}{2}}\left\|\frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\lambda_i'F^{0\prime}\hat{F}\right\|\lesssim\frac{1}{N}\sum_{i=1}^{N}\frac{\|X_i\|_F\|\lambda_i\|_F}{\sqrt{T}}\left\|\hat{\delta}\right\|$$
$$\lesssim\left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right)^{\frac{1}{2}}\left\|\hat{\delta}\right\|=O_p\left(\left\|\hat{\delta}\right\|\right),$$

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

$$T^{-\frac{1}{2}}\|I_3\|=T^{-\frac{1}{2}}\left\|\frac{1}{NT}\sum_{i=1}^{N}X_i\hat{\delta}\varepsilon_i'\hat{F}\right\|\leq\frac{1}{N}\sum_{i=1}^{N}\frac{\|X_i\|\|\varepsilon_i\|}{T}\left\|\hat{\delta}\right\|T^{-\frac{1}{2}}\left\|\hat{F}\right\|$$
$$\lesssim\left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{i=1}^{N}\|\varepsilon_i\|^2\right)^{\frac{1}{2}}\left\|\hat{\delta}\right\|=O_p\left(\left\|\hat{\delta}\right\|\right),$$

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

$$\mathbb{E}\left(\frac{1}{NT}\left\|\sum_{i=1}^{N}\lambda_i\varepsilon_i'\right\|^2\right)=\frac{1}{NT}\sum_{i,j=1}^{N}\mathbb{E}\left(\varepsilon_i'\varepsilon_j\right)E(\lambda_j'\lambda_i)\leq\frac{M}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij}=O\left(1\right)$$

by Assumption A(v), B(iv) and C(ii). Therefore $T^{-\frac{1}{2}}\left\|I_6\right\|=O_p(N^{-\frac{1}{2}})$. Analogously, $T^{-1}\left\|I_7\right\|^2=\frac{1}{T}\left\|\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\lambda_i'F^{0\prime}\hat{F}\right\|^2=O_p\left(N^{-1}\right)$. Note that the orders of 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:

$$\begin{aligned}I_{8t}&=\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_i'\hat{F}=\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{is}\hat{F}_s'\\&=\frac{1}{T}\sum_{s=1}^{T}\gamma_N\left(s,t\right)\hat{F}_s'+\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{is}-\gamma_N\left(s,t\right)\right)\hat{F}_s'\equiv I_{8t,1}+I_{8t,2},\end{aligned}$$

where $\gamma_N\left(s,t\right)=\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{is}\right)$. Then $T^{-1}\left\|I_8\right\|^2=T^{-1}\sum_{t=1}^{T}\left\|I_{8t}\right\|^2\leq 2T^{-1}\sum_{t=1}^{T}\left\|I_{8t,1}\right\|^2+2T^{-1}\sum_{t=1}^{T}\left\|I_{8t,2}\right\|^2=II_1+II_2$. By Cauchy-Schwarz inequality and Assumption C(ii),

$$II_1\leq\frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_N\left(s,t\right)^2\right)\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s'\right\|_F^2\right)\lesssim\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_N\left(s,t\right)^2=O_p\left(T^{\max(4d_\varepsilon,1)-2}\right).$$

For II_2, we have

$$\begin{aligned}II_2&=\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\xi_{st}\hat{F}_s'\right\|^2=\frac{1}{T}\frac{1}{T^2}\sum_{s,u=1}^{T}\hat{F}_s'\hat{F}_u\sum_{t=1}^{T}\xi_{st}\xi_{ut}\\&\leq\frac{1}{T}\left(\frac{1}{T^2}\sum_{s,u=1}^{T}\left(\hat{F}_s'\hat{F}_u\right)^2\right)^{\frac{1}{2}}\left(\frac{1}{T^2}\sum_{s,u=1}^{T}\left(\sum_{t=1}^{T}\xi_{st}\xi_{ut}\right)^2\right)^{\frac{1}{2}}=O_p(N^{-1}),\end{aligned}$$

where $\xi_{st}=\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{is}-\gamma_N\left(s,t\right)$, and the last equality above holds by Assumption B(iii) and the fact that

$$\begin{aligned}\frac{1}{T^2}\sum_{s,u=1}^{T}\mathbb{E}\left(\sum_{t=1}^{T}\xi_{st}\xi_{ut}\right)^2&\leq\frac{1}{T^2}\sum_{s,u=1}^{T}T^2\max_{t,v}\mathbb{E}\left|\xi_{vt}\right|^4\\&=\frac{T^2}{N^2}\max_{t,v}\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\varepsilon_{it}\varepsilon_{iv}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{iv}\right)\right]\right|^4=O\left(\frac{T^2}{N^2}\right)\end{aligned}\tag{B.1}$$

by Assumption C(iii). Then $T^{-1}\left\|I_8\right\|^2=O_p\left(N^{-1}+T^{\max(4d_\varepsilon,1)-2}\right)$. Therefore by the invertibility of H, we can conclude that $\frac{1}{T}\left\|\hat{F}-F^0H\right\|^2=O_p\left(\left\|\hat{\delta}\right\|^2+\delta_{NT}^{-2}\right)$. ■

Proof of Lemma A.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

$$\begin{aligned}\mathbb{E}\left\|\bar{A}_1\right\|^2&=\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\mathbb{E}\left[\left(\varepsilon_i'F^0F^{0\prime}\varepsilon_j\right)\left(\lambda_j'\lambda_i\right)\right]=\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{js}\right)\mathbb{E}\left(F_t^{0\prime}F_s\lambda_j'\lambda_i\right)\\&\leq\frac{\max_{i,j,t,s}\left|\mathbb{E}\left(F_t^{0\prime}F_s\lambda_j'\lambda_i\right)\right|}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\mathbb{E}\left(\varepsilon_{it}\varepsilon_{js}\right)\right|\\&\leq\frac{M}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\mathbb{E}\left(\varepsilon_{it}\varepsilon_{is}\right)\right|=O\left(N^{-1}T^{2d_\varepsilon-1}\right),\end{aligned}\tag{B.2}$$

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 A.1 and recalling that $G=\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}$, we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{\varepsilon_i'\left(\hat{F}-F^0H\right)}{T}&=\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{\varepsilon_i'\left(\hat{F}H^{-1}-F^0\right)}{T}H=\frac{1}{NT}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\left(I_1+\cdots+I_8\right)GH\\&\equiv\left(a_1+\cdots+a_8\right)GH.\end{aligned}$$

Then it remains to bound a_ℓ's by following partly the proof of Lemma A.4(ii) in Bai (2009) and using some results derived in the proof of Lemma A.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

$$\mathbb{E}\left(\frac{1}{NT}\sum_{k=1}^{N}\frac{1}{N^2}\left\|\sum_{i=1}^{N}\lambda_i\varepsilon_i'X_k\right\|^2\right)=\frac{1}{NT}\sum_{k=1}^{N}\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{js}\right)\mathbb{E}\left(\lambda_i'\lambda_jX_{kt}'X_{ks}\right)$$

$$\leq \max_{i,j,t,s} \frac{1}{N} \sum_{k=1}^{N} \left| \mathbb{E}\left(\lambda_i' \lambda_j X_{kt}' X_{ks} \right) \right| \frac{1}{N^2 T} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \left| \mathbb{E}\left(\varepsilon_{it} \varepsilon_{js} \right) \right|$$

$$\leq \frac{M}{N^2 T} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \left| \mathbb{E}\left(\varepsilon_{it} \varepsilon_{js} \right) \right| = O\left(N^{-1} T^{2d_\varepsilon} \right)$$

by Assumptions 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

$$\|a_5\| = \left\| \frac{1}{NT} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \varepsilon_i' \varepsilon_k \hat{\delta}' \frac{X_k' \hat{F}}{T} G \right\|$$

$$
\begin{aligned}
&= \left\| \frac{1}{NT} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right) \left(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \varepsilon_{kt} \hat{\delta}' \frac{X_k' \hat{F}}{T} \right) G \right\| \\
&\leq \frac{1}{NT} \sum_{t=1}^{T} \left(\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right\| \right) \left(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \| \varepsilon_{kt} \| \frac{\| X_k \|}{\sqrt{T}} \right) \left\| \hat{\delta} \right\| T^{-\frac{1}{2}} \left\| \hat{F} \right\| \| G \| \\
&\lesssim \left\| \hat{\delta} \right\| \left(\frac{1}{NT} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{NT} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \| \varepsilon_{kt} \| \frac{\| X_k \|}{\sqrt{T}} \right)^2 \right)^{\frac{1}{2}} \\
&\equiv \left\| \hat{\delta} \right\| \sqrt{a_{51}} \sqrt{a_{52}}.
\end{aligned}
$$

Note that

$$
\mathbb{E}\left(a_{51}\right) = \frac{1}{NT} \sum_{t=1}^{T} \frac{1}{N} \sum_{i,j=1}^{N} \mathbb{E}\left(\lambda_i' \lambda_j\right) \mathbb{E}\left(\varepsilon_{it} \varepsilon_{jt}\right) \leq \max_{i,j} \left| \mathbb{E}\left(\lambda_i' \lambda_j\right) \right| \frac{1}{NT} \sum_{t=1}^{T} \frac{1}{N} \sum_{i,j=1}^{N} \left| \mathbb{E}\left(\varepsilon_{it} \varepsilon_{jt}\right) \right| = O\left(N^{-1}\right)
$$

by Assumption B(iv) and C(ii), and

$$
a_{52} \leq \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{N} \sum_{k=1}^{N} \| \varepsilon_{kt} \|^2 \right) \left(\frac{1}{N} \sum_{k=1}^{N} \frac{\| X_k \|^2}{T} \right) = O_p(1)
$$

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

$$
\begin{aligned}
a_6 &= \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \left(\frac{1}{NT} \sum_{k=1}^{N} F \lambda_k \varepsilon_k' \hat{F} \right) G \\
&= \frac{1}{N^2 T^2} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \sum_{k=1}^{N} F \lambda_k \varepsilon_k' \left[F^0 H + \left(\hat{F} - F^0 H \right) \right] G \equiv a_{61} + a_{62}.
\end{aligned}
$$

Note that $\mathbb{E} \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' F \right\|^2 = \frac{1}{N^2 T^2} \sum_{i,j=1}^{N} E(\varepsilon_{it} \varepsilon_{is}) E(\lambda_i' \lambda_j F_t' F_s) = O_p\left(N^{-1} T^{2d_\varepsilon - 1}\right)$ by using the same reasoning as we analyze $\bar{A}_1$ above, we have

$$
\|a_{61}\| = \left\| \frac{1}{N^2 T^2} \sum_{i=1}^{N} \lambda_i \varepsilon_i' F \sum_{k=1}^{N} \lambda_k \varepsilon_k' F H G \right\| \lesssim \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' F \right\|^2 = O_p\left(N^{-1} T^{2d_\varepsilon - 1}\right),
$$

and

$$
\begin{aligned}
\|a_{62}\| &= \left\| \frac{1}{N^2 T^2} \sum_{i=1}^{N} \lambda_i \varepsilon_i' F \sum_{k=1}^{N} \lambda_k \varepsilon_k' \left(\hat{F} - F^0 H \right) G \right\| \\
&\lesssim \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \varepsilon_i' F \right\| \left\| \frac{1}{NT} \sum_{k=1}^{N} \lambda_k \varepsilon_k' \left(\hat{F} - F^0 H \right) \right\| \\
&\leq O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}}\right) \left(\frac{1}{N} \sum_{k=1}^{N} \| \lambda_k \|^2 \right)^{\frac{1}{2}} \left(\frac{1}{NT} \sum_{k=1}^{N} \| \varepsilon_k \|^2 \right)^{\frac{1}{2}} \frac{\left\| \hat{F} - F^0 H \right\|}{\sqrt{T}}
\end{aligned}
$$

$$= O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$$

by Lemma A.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

$$\begin{aligned}
\|a_7\| &= \left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\left(\frac{1}{NT}\sum_{k=1}^{N}\varepsilon_k\lambda_k'F^{0\prime}\hat{F}\right)G\right\| \\
&= \left\|\frac{1}{N^2T}\sum_{k=1}^{N}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\varepsilon_k\lambda_k'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\right\| \\
&\lesssim \frac{1}{NT}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{it}\right\|_F\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_{kt}\lambda_k'\right\|_F \\
&\leq \frac{1}{NT}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{it}\right\|^2 = O_p\left(N^{-1}\right)
\end{aligned}$$

using the same reasoning as above by Assumption B(iv) and C(ii).

Lastly, we study a_8 by making the following decomposition

$$a_8 = \frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_i\varepsilon_i'\sum_{k=1}^{N}\varepsilon_k\varepsilon_k'\left[F^0H+\left(\hat{F}-F^0H\right)\right]G \equiv a_{81}+a_{82}.$$

For a_{81}, we have

$$\begin{aligned}
a_{81}(HG)^{-1} &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\sum_{s=1}^{T}\varepsilon_{ks}F_s \\
&= \frac{1}{T\sqrt{N}}\frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i[\varepsilon_{it}\varepsilon_{kt}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)]\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{ks}F_s\right) \\
&\quad+\frac{1}{N\sqrt{T}}\frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{ks}F_s\right) \equiv a_{81a}+a_{81b}.
\end{aligned}$$

Note that

$$\begin{aligned}
\|a_{81a}\| &\leq \frac{1}{T\sqrt{N}}\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i[\varepsilon_{it}\varepsilon_{kt}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)]\right\|\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{ks}F_s\right\| \\
&\leq O_p\left(T^{d_\varepsilon}\right)\frac{1}{T\sqrt{N}}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i[\varepsilon_{it}\varepsilon_{kt}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)]\right\|^2\right)^{\frac{1}{2}}
\end{aligned}$$

by arguments as used in the analysis of A_1 above and Assumption C(ii). By Assumption C(iv),

$$\mathbb{E}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i[\varepsilon_{it}\varepsilon_{kt}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)]\right\|^2\right)$$

$$= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} E\left(\lambda_i'\lambda_j\right)\mathbb{E}\left\{\left[\varepsilon_{it}\varepsilon_{kt}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right]\left[\varepsilon_{js}\varepsilon_{ks}-\mathbb{E}\left(\varepsilon_{js}\varepsilon_{ks}\right)\right]\right\}$$

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

It follows that that $\|a_{81a}\| = O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon-1}\right)$. Next, noting that

$$\|a_{81b}\| \leq \frac{1}{N\sqrt{T}}\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{N}\|\lambda_i\|\left(\frac{1}{T}\sum_{t=1}^{T}\left|\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right|\right)\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{ks}F_s\right\|,$$

$$\mathbb{E}\|a_{81b}\| \leq \left(\max_k \mathbb{E}\left(\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{ks}F_s\right\|^2\right)^{\frac{1}{2}}\right)\frac{1}{N^2\sqrt{T}}\sum_{k=1}^{N}\sum_{i=1}^{N}\mathbb{E}\left(\|\lambda_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left|\left(\varepsilon_{it}\varepsilon_{kt}\right)\right|\right)$$

$$\leq MT^{d_\varepsilon}\frac{1}{N^2\sqrt{T}}\sum_{k=1}^{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left|\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right| = O\left(N^{-1}T^{d_\varepsilon-\frac{1}{2}}\right)$$

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}:

$$a_{82}G^{-1} = \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\varepsilon_i'\varepsilon_k\varepsilon_k'\left(\hat{F}-F^0H\right)$$

$$= \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{k=1}^{N}\xi_k\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T} + \frac{1}{N}\frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right)\left(\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right)$$

$$\equiv a_{82a}+a_{82b}$$

where $\xi_k = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i[\varepsilon_{it}\varepsilon_{kt}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)]$. It is easy to show that $\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\|\xi_k\|^2 = O\left(T^{2d_\varepsilon}\right)$ under Assumptions A(v), B(iv) and C(iv). Then by Lemma A.1,

$$\|a_{82a}\| = \frac{1}{\sqrt{NT}}\left\|\frac{1}{N}\sum_{k=1}^{N}\xi_k\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right\|$$

$$\leq \frac{1}{\sqrt{NT}}\left\{\frac{1}{N}\sum_{k=1}^{N}\|\xi_k\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right\|^2\right\}^{1/2}$$

$$= (NT)^{-1/2}O_p\left(T^{d_\varepsilon}\right)O_p\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right),$$

and

$$\|a_{82b}\| \leq \frac{1}{N}\left\|\frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right)\left(\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right)\right\|$$

$$\leq \frac{1}{N}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i\mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{\varepsilon_k'\left(\hat{F}-F^0H\right)}{T}\right\|^2\right\}^{1/2}$$

$$= N^{-1}O_p\left(1\right)O_p\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right).$$

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 A.2. ■

Proof of Lemma A.3. Consider the following decomposition of J_8:

$$J_8 = \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\Omega_k\hat{F}G\lambda_i + \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)\hat{F}G\lambda_i \equiv J_{81}+J_{82},$$

where $\Omega_k = \mathbb{E}\left(\varepsilon_k\varepsilon_k'\right)$ and $J_{81} = A_{NT}$. For J_{81}, we have

$$\|J_{81}\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\Omega_k\hat{F}G\lambda_i\right\| \lesssim \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|X_i\|\,\|\lambda_i\|\left\|\frac{1}{NT}\sum_{k=1}^{N}\Omega_k\right\|_{\text{sp}}$$

$$\lesssim \left(\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\sum_{t,s=1}^{T}E\left(\varepsilon_{it}\varepsilon_{is}\right)\mathbb{E}\left(\varepsilon_{kt}\varepsilon_{ks}\right)\right)^{\frac{1}{2}} \leq \left(\frac{1}{T^2}\sum_{t,s=1}^{T}|\gamma_N(s,t)|^2\right)^{\frac{1}{2}} = O_p\left(T^{\max(2d_\varepsilon,1/2)-1}\right),$$

by Assumption B(i), B(iv) and C(ii). For J_{82}, we make the decomposition

$$J_{82} = \frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)F^0HG\lambda_i$$
$$+\frac{1}{NT}\sum_{i=1}^{N}X_i'\mathbf{M}_{\hat{F}}\frac{1}{NT}\sum_{k=1}^{N}\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)(\hat{F}-F^0H)G\lambda_i \equiv J_{821}+J_{822}.$$

For J_{821}, we have

$$J_{821} = \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left[X_i'\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)F^0HG\lambda_i\right] - \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left[X_i'\frac{1}{T}\hat{F}\hat{F}'\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)F^0HG\lambda_i\right]$$
$$= \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\zeta_iF^0HG\lambda_i - \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left[X_i'\frac{1}{T}\hat{F}\hat{F}'\left(\varepsilon_k\varepsilon_k'-\Omega_k\right)F^0HG\lambda_i\right]$$
$$\equiv J_{821a}+J_{821b},$$

where $\zeta_i = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} X_{it} \left(\varepsilon_{kt}\varepsilon_{ks} - \mathbb{E}\left(\varepsilon_{kt}\varepsilon_{ks} \right) \right) F_s^{0\prime}$. Note that

$$\begin{aligned}
\mathbb{E}\left\| \zeta_i \right\|^2 &= \frac{1}{N} \sum_{i,k=1}^{N} \frac{1}{T^2} \sum_{t,s,u,v=1}^{T} \mathbb{E}\left(X_{it}' X_{iu} F_s^{0\prime} F_v^0 \right) \operatorname{cov}\left(\varepsilon_{kt}\varepsilon_{ks}, \varepsilon_{iu}\varepsilon_{iv} \right) \\
&\leq \max_{i,k,t,s,u,v} \left| \mathbb{E}\left(X_{it}' X_{iu} F_s' F_v \right) \right| \frac{1}{N} \sum_{i,k=1}^{N} \frac{1}{T^2} \sum_{t,s,u,v=1}^{T} \left| \operatorname{cov}\left(\varepsilon_{kt}\varepsilon_{ks}, \varepsilon_{iu}\varepsilon_{iv} \right) \right| = O\left(T^{4d_\varepsilon} \right),
\end{aligned}$$

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

$$\begin{aligned}
\left\| J_{821b} \right\| &\lesssim \frac{1}{\sqrt{NT}} \frac{1}{NT} \sum_{i=1}^{N} \left\| X_i' \hat{F} \right\| \left\| \lambda_i \right\| \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{F}_t \left[\varepsilon_{kt}\varepsilon_{ks} - \mathbb{E}\left(\varepsilon_{kt}\varepsilon_{ks} \right) \right] F_s^{0\prime} \right\| \\
&\lesssim \frac{1}{\sqrt{NT}} \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{F}_t \left[\varepsilon_{kt}\varepsilon_{ks} - \mathbb{E}\left(\varepsilon_{kt}\varepsilon_{ks} \right) \right] F_s^{0\prime} \right\| \\
&\lesssim \frac{1}{\sqrt{NT}} \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} F_t^0 \left[\varepsilon_{kt}\varepsilon_{ks} - \mathbb{E}\left(\varepsilon_{kt}\varepsilon_{ks} \right) \right] F_s^{0\prime} \right\| \\
&+ \frac{1}{\sqrt{NT}} \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{F}_t - H F_t^0 \right) \left[\varepsilon_{kt}\varepsilon_{ks} - \mathbb{E}\left(\varepsilon_{kt}\varepsilon_{ks} \right) \right] F_s^{0\prime} \right\| \\
&\equiv \frac{1}{\sqrt{NT}} \left\{ J_{821b1} + J_{821b2} \right\}.
\end{aligned}$$

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 A.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,

$$\left\| J_{821b2} \right\| \lesssim \frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \left(\varepsilon_k \varepsilon_k' - \Omega_k \right) \right\| = O_p\left(T^{\frac{1}{2} + d_\varepsilon} \left(\left\| \hat{\delta} \right\| + \delta_{NT}^{-1} \right) \right).$$

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

$$\begin{aligned}
\left\| J_{822} \right\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} X_i' \mathbf{M}_{\hat{F}} \frac{1}{NT} \sum_{k=1}^{N} \left(\varepsilon_k \varepsilon_k' - \Omega_k \right) \left(\hat{F} - F^0 H \right) G \lambda_i \right\| \\
&\lesssim \left\| \frac{1}{NT} \sum_{k=1}^{N} \left(\varepsilon_k \varepsilon_k' - \Omega_k \right) \right\| \frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| = O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left(\left\| \hat{\delta} \right\| + \delta_{NT}^{-1} \right) \right).
\end{aligned}$$

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 A.3. ■

Proof of Lemma A.4. Following the proof of Lemma A.8 in Bai (2009), we first study

$$\frac{1}{NT} \sum_{i=1}^{N} X_i' \left(\mathbf{M}_{F^0} - \mathbf{M}_{\hat{F}} \right) \varepsilon_i$$

by making the following decomposition

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N} X_i'\left(\mathbf{M}_{F^0}-\mathbf{M}_{\hat{F}}\right)\varepsilon_i \\
&=\frac{1}{NT}\sum_{i=1}^{N} X_i'\left(\mathbf{P}_{\hat{F}}-\mathbf{P}_{F^0}\right)\varepsilon_i \\
&=\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'\left(\hat{F}-F^0H\right)}{T}H'F'\varepsilon_i+\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'\left(\hat{F}-F^0H\right)}{T}\left(\hat{F}-F^0H\right)'\varepsilon_i \\
&+\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0H}{T}\left(\hat{F}-F^0H\right)'\varepsilon_i+\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\left[HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]F'\varepsilon_i \\
&\equiv a+b+c+d.
\end{aligned}$$

For a, we have

$$\begin{aligned}
\|a\| &=\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\hat{F}_s-H'F_s^0\right)'H'\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t^0X_{is}\varepsilon_{it}\right)\right\| \\
&\lesssim\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s-H'F_s^0\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t^0X_{is}\varepsilon_{it}\right\|^2\right)^{\frac{1}{2}} \\
&=O_p\left(N^{-1/2}T^{d_\varepsilon-1/2}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)
\end{aligned}$$

by Lemma A.1 and the fact that

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t^0X_{is}\varepsilon_{it}\right\|^2\right) \\
&=\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{r,t=1}^{T}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{jr}\right)\mathbb{E}\left(F_t^{0\prime}F_r^0X_{is}'X_{js}\right) \\
&\leq\max_{i,j,t,r,s}\mathbb{E}\left(F_t^{0\prime}F_r^0X_{is}'X_{js}\right)\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{r,t=1}^{T}\left|\mathbb{E}\left(\varepsilon_{it}\varepsilon_{jr}\right)\right|=O\left(N^{-1}T^{2d_\varepsilon-1}\right)
\end{aligned}$$

under Assumptions B(i), B(iv) and C(ii). Next, for b we have

$$\begin{aligned}
\|b\| &=\left\|\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'\left(\hat{F}-F^0H\right)}{T}\left(\hat{F}-F^0H\right)'\varepsilon_i\right\| \\
&\leq\left(\frac{1}{T}\left\|\hat{F}-F^0H\right\|^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}X_{is}\varepsilon_{it}\right\|^2\right)^{\frac{1}{2}}=O_p\left(N^{-1/2}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)
\end{aligned}$$

by Cauchy-Schwarz inequality, Lemma A.1 and the fact that

$$\mathbb{E}\left(\frac{1}{T^2}\sum_{t,s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}X_{is}\varepsilon_{it}\right\|^2\right)=\frac{1}{T^2}\sum_{s=1}^{T}\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{jt}\right)\mathbb{E}\left(X_{is}'X_{js}\right)$$

$$\leq \max_{i,j,s} \left| \mathbb{E}\left(X_{is}' X_{js}\right) \right| \frac{1}{N^2 T} \sum_{i,j=1}^{N} \sum_{t=1}^{T} \left| \mathbb{E}\left(\varepsilon_{it}\varepsilon_{jt}\right) \right| = O\left(N^{-1}\right).$$

Next, we study c by making the following decomposition:

$$c = \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\hat{F} H^{-1} - F \right)' \varepsilon_i$$
$$+ \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \left(HH' - \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right) \left(\hat{F} H^{-1} - F^0 \right)' \varepsilon_i \equiv c_1 + c_2.$$

For c_2 we have, by denoting $Q = HH' - \left(\frac{F^{0\prime} F^0}{T} \right)^{-1}$ that

$$\|c_2\| = \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} Q \left(\hat{F} H^{-1} - F^0 \right)' \varepsilon_i \right\|$$
$$= \frac{1}{NT} \sum_{i=1}^{N} \left[\varepsilon_i' \left(\hat{F} H^{-1} - F^0 \right) \otimes \left(\frac{X_i' F^0}{T} \right) \right] \mathrm{vec}\left(Q\right)$$
$$= \left[O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \right) + O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - \frac{1}{2}} \left\| \hat{\delta} \right\| \right) \right] \mathrm{vec}\left(Q\right)$$

by the proof of Lemma A.2. Next by Assumption B(iii) and Lemma A.1,

$$\frac{1}{T} F' \left(\hat{F} - F^0 H \right) = O_p\left(\left\| \hat{\delta} \right\| + \delta_{NT}^{-1} \right),$$

and the same order holds for $\frac{1}{T} \hat{F}' \left(\hat{F} - F^0 H \right)$. Then pre-multiplying $\frac{1}{T} F' \left(\hat{F} - F^0 H \right)$ by H' and using the transpose of $\frac{1}{T} \hat{F}' \left(\hat{F} - F^0 H \right)$, we can obtain

$$\mathbb{I}_R - H' \frac{F^{0\prime} F^0}{T} H = O_p\left(\left\| \hat{\delta} \right\| + \delta_{NT}^{-1} \right),$$

where the same order holds for $\mathbb{I}_R - \frac{F^{0\prime} F^0}{T} HH'$ 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 (A.2) and the proof of Lemma A.1 that

$$c_1 = \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\hat{F} H^{-1} - F^0 \right)' \varepsilon_i$$
$$= \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \left(\frac{\hat{F}' F^0}{T} \right)^{-1} \left(I_1 + \cdots + I_8 \right)' \varepsilon_i \equiv c_{1,1} + \cdots + c_{1,8}.$$

For c_{11} we have, by denoting $\tilde{G} = \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \left(\frac{\hat{F}' F^0}{T} \right)^{-1}$ that

$$\|c_{1,1}\| = \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \tilde{G} I_1' \varepsilon_i \right\| = \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \tilde{G} \frac{1}{NT} \sum_{k=1}^{N} \hat{F}' X_k \hat{\delta} \hat{\delta}' X_k' \varepsilon_i \right\|$$

$$\lesssim \left\|\hat{\delta}\right\|^2 \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} \left\| \frac{X_i' F^0}{T} \right\| \|\varepsilon_i\| \frac{1}{\sqrt{T}} \left\|\hat{F}\right\| \frac{1}{NT} \sum_{k=1}^{N} \|X_k\|^2 = O_p\left(\left\|\hat{\delta}\right\|^2\right)$$

by Assumption B(i), B(iii) and C(ii), and the fact that $\tilde{G} = O_p(1)$. For $c_{1,2}$, we have

$$\begin{aligned}
\|c_{1,2}\| &= \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \tilde{G} \frac{1}{NT} \sum_{k=1}^{N} \hat{F}' F^0 \lambda_k \hat{\delta}' X_k' \varepsilon_i \right\| \\
&\lesssim \left\|\hat{\delta}\right\| \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{X_i' F^0}{T} \right\| \frac{1}{NT} \left\| \sum_{k=1}^{N} X_k' \varepsilon_i \lambda_k' \right\| \\
&\leq \left\|\hat{\delta}\right\| \frac{1}{\sqrt{N}} \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{X_i' F^0}{T} \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{N^2 T^2} \sum_{i=1}^{N} \left\| \sum_{k=1}^{N} X_k' \varepsilon_i \lambda_k' \right\|^2 \right)^{\frac{1}{2}} \\
&= O_p\left(N^{-\frac{1}{2}} T^{d_\varepsilon - 1/2} \left\|\hat{\delta}\right\| \right),
\end{aligned}$$

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 A.1. Let ω be a $P \times 1$ nonrandom vector with $\|\omega\| = 1$.

$$\begin{aligned}
|\omega' c_{1,7}| &= \left| \frac{1}{NT} \sum_{i=1}^{N} \frac{\omega' X_i' F^0}{T} \tilde{G} \frac{1}{NT} \sum_{k=1}^{N} \hat{F}' F^0 \lambda_k \varepsilon_k' \varepsilon_i \right| \\
&= \left| \operatorname{tr}\left(\left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \frac{1}{N} \sum_{k=1}^{N} \lambda_k \varepsilon_k' \varepsilon_i \frac{\omega' X_i' F^0}{T} \right) \right| \\
&\lesssim \left\| \frac{1}{N} \sum_{k=1}^{N} \lambda_k \varepsilon_k' \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \frac{\omega' X_i'}{\sqrt{T}} \right\| \\
&\leq \frac{1}{N} \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \lambda_k \varepsilon_k' \right\| \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \varepsilon_i \frac{\omega' X_i'}{\sqrt{T}} \right\| = O_p\left(\frac{1}{N}\right),
\end{aligned}$$

where the last equality holds by the fact that

$$\mathbb{E}\left(\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \right\|^2 \right) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i,j=1}^{N} \mathbb{E}\left[\lambda_i' \lambda_j\right] \mathbb{E}\left(\varepsilon_{it} \varepsilon_{jt}\right) \leq \max_{i,j} \left| \mathbb{E}\left[\operatorname{tr}\left(\lambda_i' \lambda_j\right)\right] \right| \frac{1}{N} \sum_{i,j=1}^{N} \overline{\sigma}_{ij} = O(1)$$

by Assumption B(i), B(iii) and C(ii), and similarly $\mathbb{E}\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \varepsilon_i \omega' X_i' \right\|^2 = O_p(1)$. 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

$$\begin{aligned}
c_{1,8} = \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \tilde{G} \frac{1}{NT} \sum_{k=1}^{N} H' F^{0\prime} \varepsilon_k \varepsilon_k' \varepsilon_i + \frac{1}{NT} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \tilde{G} \frac{1}{NT} \sum_{k=1}^{N} \left(\hat{F} - F^0 H\right)' \varepsilon_k \varepsilon_k' \varepsilon_i \\
\equiv c_{1,81} + c_{1,82}.
\end{aligned}$$

Note that

$$c_{1,81} = \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}H'F^{0\prime}\varepsilon_k[\varepsilon_k'\varepsilon_i - \mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)] + \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}H'F^{0\prime}\varepsilon_k\mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)$$
$$\equiv c_{1,811} + c_{1,812}.$$

For $c_{1,811}$,

$$\begin{aligned}
\left|\omega' c_{1,811}\right| &= \left\| \frac{1}{N^2T^3}\sum_{i,k=1}^{N} F^0\tilde{G}H'F^{0\prime}\varepsilon_k[\varepsilon_k'\varepsilon_i - \mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i' \right\| \\
&\lesssim \left\| \frac{1}{N^2T^{5/2}}\sum_{i,k=1}^{N} F^{0\prime}\varepsilon_k[\varepsilon_k'\varepsilon_i - \mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i' \right\| \\
&\lesssim \frac{1}{T}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{NT}\sum_{i=1}^{N}[\varepsilon_k'\varepsilon_i - \mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i'\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|F^{0\prime}\varepsilon_k\right\|^2\right)^{\frac{1}{2}} \\
&= \frac{1}{T}O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon}\right)O_p\left(T^{d_\varepsilon}\right) = O_p\left(N^{-\frac{1}{2}}T^{2d_\varepsilon - 1}\right),
\end{aligned}$$

where the last equality holds by the fact that

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{NT}\sum_{i=1}^{N}[\varepsilon_k'\varepsilon_i - \mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)]\omega' X_i'\right\|^2\right) \\
&= \frac{1}{N^3T}\sum_{i,j,k=1}^{N}\mathbb{E}\left[\frac{\omega' X_i' X_j\omega}{T}\right]\sum_{t,s=1}^{T}\mathbb{E}\left\{\left[\varepsilon_{it}\varepsilon_{kt} - \mathbb{E}\left(\varepsilon_{it}\varepsilon_{kt}\right)\right]\left[\varepsilon_{js}\varepsilon_{ks} - \mathbb{E}\left(\varepsilon_{js}\varepsilon_{ks}\right)\right]\right\} \\
&\le \max_{i,j}\mathbb{E}\left[\frac{\omega' X_i' X_j\omega}{T}\right]\frac{1}{N^3T}\sum_{i,j,k=1}^{N}\sum_{t,s=1}^{T}\mathbb{E}\left[\left(\varepsilon_{kt}\varepsilon_{it} - \mathbb{E}\left(\varepsilon_{kt}\varepsilon_{it}\right)\right)\left(\varepsilon_{ks}\varepsilon_{js} - \mathbb{E}\left(\varepsilon_{ks}\varepsilon_{js}\right)\right)\right] \\
&\lesssim \frac{1}{N^3T}\sum_{i,j,k=1}^{N}\sum_{t,s=1}^{T}\left|\mathrm{cov}\left(\varepsilon_{it}\varepsilon_{kt}, \varepsilon_{js}\varepsilon_{ks}\right)\right| = O\left(N^{-1}T^{2d_\varepsilon}\right) \qquad \text{(B.3)}
\end{aligned}$$

by Assumption B(i), B(iii) and C(iv). Next,

$$\left\|c_{1,812}\right\| = \left\|\frac{1}{NT}\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\tilde{G}\sum_{k=1}^{N}H'F^{0\prime}\varepsilon_k\frac{\mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)}{T}\right\| \lesssim \frac{1}{NT}\frac{1}{N}\sum_{i,k=1}^{N}\left\|\frac{X_i'F^0}{T}\right\|_F\left\|F^{0\prime}\varepsilon_k\right\|\overline{\sigma}_{ik},$$

where the expectation of the term is bounded above by

$$\begin{aligned}
&\frac{1}{NT}\max_{i,k}\mathbb{E}\left(\left\|\frac{X_i'F^0}{T}\right\|\left\|F^{0\prime}\varepsilon_k\right\|\right)\frac{1}{N}\sum_{i,k=1}^{N}\overline{\sigma}_{ik} \\
&\le \frac{M}{N}\max_i\left(\mathbb{E}\left\|\frac{X_i'F^0}{T}\right\|^2\right)^{\frac{1}{2}}\max_k\left(\mathbb{E}\left\|\frac{F^{0\prime}\varepsilon_k}{T}\right\|^2\right)^{\frac{1}{2}} = O\left(N^{-1}T^{d_\varepsilon - \frac{1}{2}}\right).
\end{aligned}$$

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

$$c_{1,82} = \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\left(\hat{F}-F^0H\right)'\varepsilon_k\left[\varepsilon_k'\varepsilon_i - \mathbb{E}\left(\varepsilon_k'\varepsilon_i\right)\right]$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\tilde{G}\frac{1}{NT}\sum_{k=1}^{N}\left(\hat{F}-F^0H\right)'\varepsilon_k\mathbb{E}\left(\varepsilon_k'\varepsilon_i\right) \equiv c_{1,821}+c_{1,822}.$$

For the first term on the r.h.s., we have

$$\|c_{1,821}\| \lesssim \frac{1}{T}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{NT}\sum_{i=1}^{N}X_i'\sum_{t=1}^{T}\left[\varepsilon_{kt}\varepsilon_{it}-\mathbb{E}\left(\varepsilon_{kt}\varepsilon_{it}\right)\right]\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\left(\hat{F}-F^0H\right)'\varepsilon_k\right\|^2\right)^{\frac{1}{2}}$$
$$= T^{-1}O\left(N^{-1/2}T^{d_\varepsilon}\right)O_p\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right) = O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-1}\left\|\hat{\delta}\right\|+N^{-1/2}T^{d_\varepsilon-1}\delta_{NT}^{-1}\right)$$

by (B.3) and the derivation of order of the term A_2 in the proof of Lemma A.2. In addition,

$$\|c_{1,822}\| \lesssim \frac{1}{NT}\frac{1}{N}\sum_{i,k=1}^{N}\left\|\frac{X_i'F^0}{T}\right\|\left\|\left(\hat{F}-F^0H\right)'\varepsilon_k\right\|\overline{\sigma}_{ik} = O_p\left(N^{-1}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)$$

by Lemma A.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.

$$\begin{aligned}\|d\| &= \left\|\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'F^0}{T}QF'\varepsilon_i\right\| \leq \frac{1}{T}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{X_i'F^0}{T}\right\|\left\|F^{0\prime}\varepsilon_i\right\|\|Q\| \\ &\lesssim \frac{1}{T}\|Q\| = O_p\left(T^{-1}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)\right)\end{aligned}$$

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,

$$\frac{1}{NT}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\right)'\left(\mathbf{M}_{F^0}-\mathbf{M}_{\hat{F}}\right)\varepsilon_i \equiv \frac{1}{NT}\sum_{i=1}^{N}V_i'\left(\mathbf{M}_{F^0}-\mathbf{M}_{\hat{F}}\right)\varepsilon_i,$$

can be expressed by replacing X_i with V_i, and apply the same arguments and probability order as above. Then we concludes that

$$\frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{\hat{F}}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{\hat{F}}\right]\varepsilon_i$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\left[X_i'\mathbf{M}_{F^0}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'\mathbf{M}_{F^0}\right]\varepsilon_i - C_{NT}$$
$$+O_p\left(N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\left(\left\|\hat{\delta}\right\|+\delta_{NT}^{-1}\right)+\left\|\hat{\delta}\right\|^2+N^{-\frac{1}{2}}T^{2d_\varepsilon-1}\right).$$

This completes the proof of Lemma A.4. ■

Proof of Lemma A.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 $X_{p,it}$ denotes 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 $\hat{F}_{X_{p,i}}\left(\gamma_L\right)$ and $\hat{F}_{\varepsilon,i}\left(\gamma_L\right)$ 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 (1994) to obtain

$$\frac{\hat{F}_{X_{p,i}}\left(\gamma_L\right)}{F_{X_{p,i}}\left(\gamma_L\right)}\xrightarrow{p}1\text{ and }\frac{\hat{F}_{\varepsilon,i}\left(\gamma_L\right)}{F_{\varepsilon,i}\left(\gamma_L\right)}\xrightarrow{p}1\text{ as }T\to\infty,\tag{B.4}$$

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 $\Upsilon_{i,XX,pp}$ denotes 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,

$$\begin{aligned}\left|\frac{1}{NT}\sum_{i=1}^{N}W_{X_p,i}^{*}W_{\varepsilon,i}\right|&\lesssim\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\Upsilon_{i,XX,pp}}{1-2d_{X_p}}\gamma_L^{1-2d_{X_p}}\right|^{\frac{1}{2}}\left|\frac{\Upsilon_{i,\varepsilon\varepsilon}}{1-2d_{\varepsilon}}\gamma_L^{1-2d_{\varepsilon}}\right|^{\frac{1}{2}}\\&=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\Upsilon_{i,XX,pp}\Upsilon_{i,\varepsilon\varepsilon}}{\left(1-2d_{X_p}\right)\left(1-2d_{\varepsilon}\right)}\right)^{\frac{1}{2}}\gamma_L^{1-d_{X_p}-d_{\varepsilon}}=o_p(1)\end{aligned}$$

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). 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 $\left|\frac{1}{T}\bar{W}_{F_r}^{*}W_{\varepsilon,i}\right|=O_p\left(\gamma_L^{\frac{1}{2}-d_{\varepsilon}}\right)$ 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

$$\sup_{\tilde{W}_F\in\mathcal{W}}\left\|\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}^{*}\mathbf{M}_{\tilde{W}_F}W_{\varepsilon,i}\right\|=o_p(1).$$

(ii) and (iii): The proof is similar to that of (i) and thus omitted. ■

Proof of Lemma A.6. Let $\tilde{\delta}=\tilde{\beta}-\beta$. As in the proof of Lemma A.1, we consider the following eigenvalue problem

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)\left(W_{Y,i}-W_{X,i}\tilde{\beta}\right)^{*}\right]\hat{W}_F=\hat{W}_FV_{NL}. \tag{B.5}$$

By expanding $W_{Y,i}$ in (B.5), we have

$$\begin{aligned}\hat{W}_FV_{NL}=&\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}\tilde{\delta}\tilde{\delta}'W_{X,i}^{*}\hat{W}_F-\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}\tilde{\delta}\lambda_i'W_{F^0}^{*}\hat{W}_F-\frac{1}{NT}\sum_{i=1}^{N}W_{X,i}\tilde{\delta}W_{\varepsilon,i}^{*}\hat{W}_F\\&-\frac{1}{NT}\sum_{i=1}^{N}W_{F^0}\lambda_i\tilde{\delta}'W_{X,i}^{*}\hat{W}_F-\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}\tilde{\delta}'W_{X,i}^{*}\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}W_{F^0}\lambda_iW_{\varepsilon,i}^{*}\hat{W}_F\\&+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}\lambda_i'W_{F^0}^{*}\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}W_{\varepsilon,i}W_{\varepsilon,i}^{*}\hat{W}_F+\frac{1}{NT}\sum_{i=1}^{N}\tilde{W}_{F^0}\tilde{\lambda}_i\tilde{\lambda}_i'\tilde{W}_{F^0}^{*}\hat{W}_F\\&\equiv\tilde{I}_1+\cdots+\tilde{I}_9.\end{aligned} \tag{B.6}$$

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 $\hat{W}_FV_{NL}-\tilde{W}_{F^0}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)\left(\tilde{W}_{F^0}^{*}\hat{W}_F/T\right)=\tilde{I}_1+\cdots+\tilde{I}_8$. 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

$$T^{-\frac{1}{2}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|=T^{-\frac{1}{2}}\left\|\left[\hat{W}_FV_{NL}\left(\tilde{W}_{F^0}^{*}\hat{W}_F/T\right)^{-1}\left(\tilde{\Lambda}'\tilde{\Lambda}/N\right)^{-1}-\tilde{W}_{F^0}\right]\tilde{H}\right\|$$

$$
\begin{aligned}
&= T^{-\frac{1}{2}} \left\| \left(\tilde{I}_1 + \cdots + \tilde{I}_8\right) \left(\tilde{W}^*_{F^0} \hat{W}_F / T\right)^{-1} \left(\tilde{\Lambda}' \tilde{\Lambda} / N\right)^{-1} \tilde{H} \right\| \\
&\lesssim T^{-\frac{1}{2}} \left(\left\| \tilde{I}_1 \right\| + \cdots \left\| \tilde{I}_8 \right\| \right) \| V_{NL}^{-1} \|
\end{aligned}
$$

given the invertibility of $\tilde{W}^*_{F^0} \hat{W}_F / T$ by Proposition 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

$$
\begin{aligned}
T^{-\frac{1}{2}} \left\| \tilde{I}_1 \right\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W^*_{X,i} \tilde{\delta} \tilde{\delta}' W^*_{X,i} \hat{W}_F \right\| \\
&\leq \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \| W_{X,i} \|^2 \left\| \tilde{\delta} \right\|^2 T^{-\frac{1}{2}} \left\| \hat{W}_F \right\| \lesssim \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \| W_{X,i} \|^2 \left\| \tilde{\delta} \right\|^2 \\
&= O_p\left(\gamma_L^{1-2d_{X,\max}} \right) \left\| \tilde{\delta} \right\|^2 = o_p\left(\gamma_L^{1-2d_{X,\max}} \left\| \tilde{\delta} \right\| \right),
\end{aligned}
$$

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 A.5 under Assumption B*(i). Similarly,

$$
\begin{aligned}
T^{-\frac{1}{2}} \left\| \tilde{I}_2 \right\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{X,i} \tilde{\delta} \tilde{\lambda}_i' \tilde{W}^*_{F^0} \hat{W}_F \right\| \lesssim \frac{1}{NT^{\frac{1}{2}}} \sum_{i=1}^{N} \| W_{X,i} \| \left\| \tilde{\lambda}_i \right\| T^{-\frac{1}{2}} \left\| \tilde{W}_{F^0} \right\| \left\| \tilde{\delta} \right\| \\
&\lesssim \left\{ \frac{1}{NT} \sum_{i=1}^{N} \| W_{X,i} \|^2 \right\}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{\lambda}_i \right\|^2 \right\}^{1/2} \left\| \tilde{\delta} \right\| \\
&= O_p\left(\gamma_L^{1/2-d_{X,\max}} \right) O_p\left(\gamma_L^{1/2-d_{F,\max}} \right) \left\| \tilde{\delta} \right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_{F,\max}} \left\| \tilde{\delta} \right\| \right)
\end{aligned}
$$

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 A.5. Analogously, we have

$$
\begin{aligned}
T^{-\frac{1}{2}} \left\| \tilde{I}_3 \right\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{X,i} \tilde{\delta} W^*_{\varepsilon,i} \hat{W}_F \right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon} \left\| \tilde{\delta} \right\| \right), \\
T^{-\frac{1}{2}} \left\| \tilde{I}_4 \right\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \tilde{W}_{F^0} \tilde{\lambda}_i \tilde{\delta}' W^*_{X,i} \hat{W}_F \right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_F} \left\| \tilde{\delta} \right\| \right), \text{ and} \\
T^{-\frac{1}{2}} \left\| \tilde{I}_5 \right\| &= T^{-\frac{1}{2}} \left\| \frac{1}{NT} \sum_{i=1}^{N} W_{\varepsilon,i} \tilde{\delta}' W^*_{X,i} \hat{W}_F \right\| = O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon} \left\| \tilde{\delta} \right\| \right).
\end{aligned}
$$

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 $T^{-\frac{1}{2}}\left\|\tilde{I}_7\right\| = T^{-\frac{1}{2}}\left\|\tilde{I}_6\right\|$. For $\tilde{I}_8$, we follow the reasoning as used in the proof of Lemma A.1 and consider the transpose of the l-th row of $\tilde{I}_8$ as

$$\tilde{I}_{8,l} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{k=1}^{L}\mathbb{E}\left(W_{\varepsilon,il}W_{\varepsilon,ik}^*\right)\hat{W}_{F,k} + \frac{1}{T}\sum_{k=1}^{L}\left(\frac{1}{N}\sum_{i=1}^{N}W_{\varepsilon,il}W_{\varepsilon,ik}^* - \gamma_N^W(k,l)\right)\hat{W}_{F,k} \equiv \tilde{I}_{8,l1} + \tilde{I}_{8,l2},$$

where $\gamma_N^W(k,l) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(W_{\varepsilon,il}W_{\varepsilon,ik}^*\right)$ and $\hat{W}_{F,k}$ denotes the k-th column of $\hat{W}_F$. Note that

$$\begin{aligned}
\frac{1}{T}\sum_{l=1}^{L}\left\|\tilde{I}_{8,l1}\right\|^2 &= \frac{1}{N^2T^2}\sum_{l=1}^{L}\left\|\sum_{i=1}^{N}\sum_{k=1}^{L}\mathbb{E}\left(W_{\varepsilon,il}W_{\varepsilon,ik}^*\right)\hat{W}_{F,k}\right\|^2 \\
&= \frac{1}{N^2T^3}\sum_{l=1}^{L}\sum_{i,j=1}^{N}\sum_{k,m=1}^{L}\mathbb{E}\left(W_{\varepsilon,il}W_{\varepsilon,ik}^*\right)\mathbb{E}\left(W_{\varepsilon,jl}^*W_{\varepsilon,jm}\right)\hat{W}_{F,k}^*\hat{W}_{F,m} \\
&\le \frac{1}{N^2T^3}\sum_{l=1}^{L}\sum_{i,j=1}^{N}\sum_{k,m=1}^{L}\sqrt{\mathbb{E}\left|W_{\varepsilon,il}W_{\varepsilon,ik}^*\right|^2\mathbb{E}\left|W_{\varepsilon,jl}^*W_{\varepsilon,jm}\right|^2}\left\|\hat{W}_{F,k}\right\|\left\|\hat{W}_{F,m}\right\| \\
&\le \frac{1}{N^2T^3}\sum_{l=1}^{L}\sum_{i,j=1}^{N}\sum_{k,m=1}^{L}\gamma_l^{-2d_\varepsilon}\gamma_k^{-d_\varepsilon}\gamma_m^{-d_\varepsilon}\overline{\sigma}_{ij}^W\left\|\hat{W}_{F,k}\right\|\left\|\hat{W}_{F,m}\right\| \\
&\lesssim \frac{1}{NT^3}\left(\sum_{l=1}^{L}\gamma_l^{-2d_\varepsilon}\right)^2\left(\sum_{l=k}^{L}\left\|\hat{W}_{F,k}\right\|^2\right) = O_p\left(\frac{1}{N}\gamma_L^{2-4d_\varepsilon}\right)
\end{aligned}$$

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). Then $T^{-\frac{1}{2}}\left\|\tilde{I}_8\right\| = O_p\left(N^{-1/2}\gamma_L^{1-2d_\varepsilon}\right)$. In sum, we have

$$\begin{aligned}
T^{-\frac{1}{2}}\left\|\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right\| &= O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon} + N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right) \\
&= O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right),
\end{aligned}$$

where $\delta_{W1,NT} = \gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$. ■

Proof of Lemma A.7. The proof of this lemma parallels that of Lemma A.2. Note that

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^*\hat{W}_F}{T} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^*\tilde{W}_{F^0}\tilde{H}}{T} + \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)}{T} \equiv A_1 + A_2. \tag{B.7}$$

For A_1, we have

$$\begin{aligned}
\|A_1\| &\le \left\|\tilde{H}\right\|\frac{1}{NT}\left\|\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^*\tilde{W}_{F^0}\right\| \\
&= O_p\left(\gamma_L^{1-2d_{F,\max}}\right)O_p\left(N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right) = O_p\left(N^{-\frac{1}{2}}N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right)
\end{aligned}$$

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 A.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 A.6, we have for A_2 that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)}{T} &= \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\left(\hat{W}_F\tilde{H}^{-1}-\tilde{W}_{F^0}\right)}{T}\tilde{H}\\
&= \frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^{*}\left(\tilde{I}_1+\cdots+\tilde{I}_8\right)\tilde{G}}{T}\tilde{H}\\
&\lesssim \frac{1}{NT}\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\left(\tilde{I}_1+\cdots+\tilde{I}_8\right) \equiv a_1+\cdots+a_8,
\end{aligned}$$

using the fact that $\left\|\tilde{G}\tilde{H}\right\| = \left\|V_{NL}^{-1}\right\| = O_p(1)$. For a_1, we have

$$\begin{aligned}
\|a_1\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}\tilde{\delta}\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right\|\\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\frac{1}{NT}\sum_{k=1}^{N}\left\|W_{X,k}^{*}\right\|^2\left\|\tilde{\delta}\right\|^2\\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1-2d_{X,\max}}\right)\left\|\tilde{\delta}\right\|^2 = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{X,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|^2\right)
\end{aligned}$$

by Assumption B*(i) and C*(i). Similarly,

$$\begin{aligned}
\|a_2\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}\tilde{\delta}\lambda_k'W_{F^0}^{*}\hat{W}_F\right\|\\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\left\|W_{X,k}\right\|\left\|\lambda_k\right\|T^{-\frac{1}{2}}\left\|W_{F^0}\right\|\left\|\tilde{\delta}\right\|\\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\left\{\frac{1}{NT}\sum_{k=1}^{N}\left\|W_{X,k}\right\|^2\right\}^{1/2}T^{-\frac{1}{2}}\left\|W_{F^0}\right\|\left\|\tilde{\delta}\right\|\\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1/2-d_{X,\max}}\right)O_p\left(\gamma_L^{1/2-d_{F,\max}}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{F,\max}-d_{X,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|\right),
\end{aligned}$$

$$\begin{aligned}
\|a_3\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{X,k}\tilde{\delta}W_{\varepsilon,k}^{*}\hat{W}_F\right\|\\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_i W_{\varepsilon,i}^{*}\right\|\left\{\frac{1}{NT}\sum_{k=1}^{N}\left\|W_{X,k}\right\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{k=1}^{N}\left\|W_{\varepsilon,k}\right\|^2\right\}^{1/2}\left\|\tilde{\delta}\right\|
\end{aligned}$$

$$= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right) O_p\left(\gamma_L^{1/2-d_{X,\max}}\right) O_p\left(\gamma_L^{1/2-d_\varepsilon}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-2d_\varepsilon}\left\|\tilde{\delta}\right\|\right),$$

$$\begin{aligned}
\|a_4\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{F^0}\lambda_k\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}\right\|\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\|W_{X,k}\|\,\|\lambda_k\|\,T^{-\frac{1}{2}}\|W_{F^0}\|\left\|\tilde{\delta}\right\| \\
&\leq \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}\right\|\left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2\right\}^{1/2}T^{-\frac{1}{2}}\|W_{F^0}\|\left\|\tilde{\delta}\right\| \\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1/2-d_{X,\max}}\right)O_p\left(\gamma_L^{1/2-d_{F,\max}}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-d_{F,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|\right),
\end{aligned}$$

and

$$\begin{aligned}
\|a_5\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{1}{T}W_{\varepsilon,i}^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}\tilde{\delta}'W_{X,k}^{*}\hat{W}_F\right\| \\
&\lesssim \frac{1}{NT^{\frac{1}{2}}}\left\|\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}\right\|\left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{\varepsilon,k}\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{k=1}^{N}\|W_{X,k}\|^2\right\}^{1/2}\left\|\tilde{\delta}\right\| \\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)O_p\left(\gamma_L^{1/2-d_\varepsilon}\right)O_p\left(\gamma_L^{1/2-d_{X,\max}}\right)\left\|\tilde{\delta}\right\| = O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-2d_\varepsilon}\left\|\tilde{\delta}\right\|\right).
\end{aligned}$$

For a_6, we have $a_6 = \frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}\sum_{k=1}^{N}W_{F^0}\lambda_kW_{\varepsilon,k}^{*}\left[\tilde{W}_{F^0}\tilde{H}+\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\right] \equiv a_{6,1}+a_{6,2}$. Note that

$$\begin{aligned}
\|a_{6,1}\| &= \left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}W_{F^0}\sum_{k=1}^{N}\lambda_kW_{\varepsilon,k}^{*}\tilde{W}_{F^0}\tilde{H}\right\| \leq \left\|\tilde{\Gamma}_F\tilde{H}\right\|\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}W_{F^0}\right\|^2 \\
&\lesssim \gamma_L^{\frac{1}{2}-d_{F,\max}}\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}W_{F^0}\right\|^2 \\
&= O_p\left(\gamma_L^{\frac{1}{2}-d_{F,\max}}\right)O_p\left(N^{-1}L^{-1}\gamma_L^{2-2d_\varepsilon}\right)O_p\left(\gamma_L^{-2d_{F,\max}}\right) = O_p\left(N^{-1}L^{-1}\gamma_L^{\frac{5}{2}-3d_{F,\max}-2d_\varepsilon}\right),
\end{aligned}$$

by Assumption D*(iii). Next

$$\begin{aligned}
\|a_{6,2}\| &= \left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}W_{F^0}\sum_{k=1}^{N}\lambda_kW_{\varepsilon,k}^{*}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\right\| \\
&\leq \frac{1}{N^2T}\left\|\sum_{i=1}^{N}\lambda_iW_{\varepsilon,i}^{*}\right\|^2T^{-\frac{1}{2}}\|W_{F^0}\|\,T^{-\frac{1}{2}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\| \\
&= O_p\left(N^{-1}\gamma_L^{1-2d_\varepsilon}\right)O_p\left(\gamma_L^{\frac{1}{2}-d_{F,\max}}\right)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right) \\
&= O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right),
\end{aligned}$$

by Lemma A.1, where $\delta_{W1,NT} = \gamma_L^{\frac{1}{2}-d_{X,\max}}\left(\gamma_L^{\frac{1}{2}-d_{F,\max}} + \gamma_L^{\frac{1}{2}-d_\varepsilon}\right)$. Note that the order of $\left\|\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^N \lambda_i W_{\varepsilon,i}^*\right\|^2$ is obtained because

$$\mathbb{E}\left\|\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^N \lambda_i W_{\varepsilon,i}^*\right\|^2 \frac{1}{T} = \frac{1}{N^2T}\sum_{i,k=1}^N \mathbb{E}\left(\lambda_k'\lambda_i\right)\mathbb{E}\left(W_{\varepsilon,i}^* W_{\varepsilon,k}\right)$$

$$\leq \max_{i,k}\left|\mathbb{E}\left(\lambda_k'\lambda_i\right)\right|\frac{1}{N^2T}\sum_{i,k=1}^N\sum_{l=1}^L\left|\mathbb{E}\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right| = O\left(N^{-1}\gamma_L^{1-2d_\varepsilon}\right)$$

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

$$\begin{aligned}\|a_7\| &= \left\|\frac{1}{N}\sum_{i=1}^N \lambda_i \frac{1}{T}W_{\varepsilon,i}^*\left(\frac{1}{NT}\sum_{k=1}^N W_{\varepsilon,k}\lambda_k' W_{F^0}^*\hat{W}_F\right)\right\| \\ &\lesssim \frac{1}{N^2T}\left\|\sum_{i=1}^N \lambda_i W_{\varepsilon,i}^*\right\|^2 T^{-\frac{1}{2}}\left\|W_{F^0}\right\| \\ &= O_p\left(N^{-1}\gamma_L^{1-2d_\varepsilon}\right)O_p\left(\gamma_L^{\frac{1}{2}-d_{F,\max}}\right) = O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\right).\end{aligned}$$

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

$$\begin{aligned}a_{8,1} &= \frac{1}{N^2T^2}\sum_{i=1}^N\sum_{k=1}^N \lambda_i\sum_{l=1}^L\left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - \mathbb{E}\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right]W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H} \\ &+ \frac{1}{N^2T^2}\sum_{i=1}^N\sum_{k=1}^N \lambda_i\sum_{l=1}^L \mathbb{E}\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H} \equiv a_{8,11} + a_{8,12}.\end{aligned}$$

For $a_{8,11}$, we have

$$\begin{aligned}\|a_{8,11}\| &\leq \frac{1}{N^2T^2}\sum_{k=1}^N\left\|\sum_{i=1}^N \lambda_i\sum_{l=1}^L\left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - \mathbb{E}\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right]\right\|\left\|W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H}\right\| \\ &\leq \frac{1}{\sqrt{N}T}\left(\frac{1}{N^2}\sum_{k=1}^N\left\|\sum_{i=1}^N \lambda_i\sum_{l=1}^L\left[W_{\varepsilon,il}^* W_{\varepsilon,kl} - \mathbb{E}\left(W_{\varepsilon,il}^* W_{\varepsilon,kl}\right)\right]\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT^2}\sum_{k=1}^N\left\|W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H}\right\|^2\right)^{\frac{1}{2}} \\ &= O\left(N^{-\frac{1}{2}}T^{-1}\right)O_p\left(L\gamma_L^{-2d_\varepsilon}\right)O_p\left(L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right)\end{aligned}$$

$$= O_p\left(N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{5}{2}-2d_{F,\max}-3d_\varepsilon}\right)$$

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 A.2. To be specific,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{N^2}\sum_{k=1}^{N}\left\|\sum_{i=1}^{N}\lambda_i\sum_{l=1}^{L}\left[W^*_{\varepsilon,il}W_{\varepsilon,kl}-\mathbb{E}\left(W^*_{\varepsilon,il}W_{\varepsilon,kl}\right)\right]\right\|^2\right]\\
&=\frac{1}{N^2}\sum_{h,i,k=1}^{N}\mathbb{E}\left(\lambda_i^*\lambda_h\right)\sum_{l,m=1}^{L}\mathbb{E}\left\{\left[W^*_{\varepsilon,il}W_{\varepsilon,kl}-\mathbb{E}\left(W^*_{\varepsilon,il}W_{\varepsilon,kl}\right)\right]\left[W^*_{\varepsilon,hm}W_{\varepsilon,km}-\mathbb{E}\left(W^*_{\varepsilon,hm}W_{\varepsilon,km}\right)\right]\right\}\\
&=\frac{1}{N^2}\sum_{h,i,k=1}^{N}\mathbb{E}\left(\lambda_i^*\lambda_h\right)\sum_{l,m=1}^{L}\operatorname{cov}\left[W^*_{\varepsilon,il}W_{\varepsilon,kl},W^*_{\varepsilon,hm}W_{\varepsilon,km}\right]\\
&\le\frac{\max_{i,h}\mathbb{E}\left(\lambda_i^*\lambda_h\right)}{N^2}\sum_{h,i,k=1}^{N}\sum_{l,m=1}^{L}\left|\operatorname{cov}\left[W^*_{\varepsilon,il}W_{\varepsilon,kl},W^*_{\varepsilon,hm}W_{\varepsilon,km}\right]\right|=O\left(L^2\gamma_L^{-4d_\varepsilon}\right).
\end{aligned}$$

Next, $a_{8,12}$ follows that

$$\begin{aligned}
\|a_{8,12}\|&\le\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\|\lambda_i\|\left(\sum_{l=1}^{L}\left|\mathbb{E}\left(W^*_{\varepsilon,il}W_{\varepsilon,kl}\right)\right|\right)\left\|W^*_{\varepsilon,k}\tilde{W}_{F^0}\tilde{H}\right\|\\
&\lesssim\gamma_L^{1-2d_{F,\max}}\frac{1}{N^2T}\sum_{i=1}^{N}\|\lambda_i\|\sum_{k=1}^{N}\left(\sum_{l=1}^{L}\left|\mathbb{E}\left(W^*_{\varepsilon,il}W_{\varepsilon,kl}\right)\right|\right)\frac{1}{\sqrt{T}}\left\|W_{\varepsilon,k}\tilde{W}_{F^0}\right\|\\
&\lesssim\frac{\gamma_L^{1-2d_{F,\max}}}{\sqrt{N}T}\left(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\sum_{l=1}^{L}\left|\mathbb{E}\left(W^*_{\varepsilon,il}W_{\varepsilon,kl}\right)\right|\right)^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{k=1}^{N}\frac{1}{T}\left\|W_{\varepsilon,k}\tilde{W}_{F^0}\right\|^2\right)^{\frac{1}{2}}\\
&=O_p\left(\gamma_L^{1-2d_{F,\max}}N^{-\frac{1}{2}}T^{-1}\right)O_p(1)\,O\left(L\gamma_L^{-2d_\varepsilon}\right)O_p\left(L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)\\
&=O_p\left(N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{5}{2}-2d_{F,\max}-3d_\varepsilon}\right)
\end{aligned}$$

by Cauchy-Schwarz inequality and Assumptions B(iv), C*(i), C*(ii) and D*(iii). To be specific,

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\sum_{l=1}^{L}\left|\mathbb{E}\left(W^*_{\varepsilon,il}W_{\varepsilon,kl}\right)\right|\right)^2&\le\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\sum_{l=1}^{L}\sqrt{\mathbb{E}\left(W^*_{\varepsilon,il}W_{\varepsilon,kl}\right)^2}\right)^2\\
&\le\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\overline{\sigma}_{ij}^{W}\right)^2\left(\sum_{l=1}^{L}\gamma_l^{-2d_\varepsilon}\right)^2=O\left(L^2\gamma_L^{-4d_\varepsilon}\right)
\end{aligned}$$

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 $T^{-\frac{1}{2}}\left\|\hat{W}_F-\tilde{W}_{F^0}H\right\|$ and obtain

$$\|a_{8,2}\|=O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-3d_\varepsilon}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right),$$

where $\delta_{W1,NT}$ is defined in Lemma A.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

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i\frac{W_{\varepsilon,i}^*\hat{W}_F}{T}\right\| &= O_p\left(N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right) + O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{X,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|^2\right)\\
&+ O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-d_{F,\max}-d_\varepsilon}\left\|\tilde{\delta}\right\|\right) + O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-d_{X,\max}-2d_\varepsilon}\left\|\tilde{\delta}\right\|\right)\\
&+ O_p\left(N^{-1}L^{-1}\gamma_L^{\frac{5}{2}-3d_{F,\max}-2d_\varepsilon}\right) + O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\right)\\
&+ O_p\left(N^{-1}\gamma_L^{\frac{3}{2}-d_{F,\max}-2d_\varepsilon}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)\\
&+ O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-3d_\varepsilon}\left(L^{-\frac{1}{2}}\gamma_L^{1-2d_{F,\max}} + \delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right)\\
&= O_p\left(\delta_{W,NL}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_{F,\max}-d_\varepsilon}\right),\end{aligned}$$

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 A.7. ■

Proof of Lemma A.8. For $\tilde{J}_8$ we have the following decomposition:

$$\begin{aligned}\tilde{J}_8 &= -\frac{1}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\right)\\
&= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\right) + \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\hat{W}_F^*W_{\varepsilon,k}W_{\varepsilon,k}^*\hat{W}_F\check{G}\lambda_i\right)\\
&\equiv \tilde{J}_{8,a} + \tilde{J}_{8,b}.\end{aligned}$$

First, we can decompose $\tilde{J}_{8,a}$ as follows

$$\begin{aligned}\tilde{J}_{8,a} &= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\tilde{W}_{F^0}\tilde{H}\check{G}\lambda_i\right) - \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)\\
&= -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\sum_{j=1}^{L}\overline{W}_{X,ij}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}_{F^0,j}'\tilde{H}\check{G}\lambda_i\right) - \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(\sum_{j\neq l}^{L}\overline{W}_{X,ij}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kl}\tilde{W}_{F^0,l}'\tilde{H}\check{G}\lambda_i\right)\\
&- \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\mathrm{Re}\left(W_{X,i}^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F - \tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)\\
&\equiv \tilde{J}_{8,a1} + \tilde{J}_{8,a2} + \tilde{J}_{8,a3},\end{aligned}$$

and $\tilde{J}_{8,b}$ similarly

$$\tilde{J}_{8,b} = \frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\tilde{H}^*\sum_{j=1}^{L}\overline{\overline{W}}_{F^0,j}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}'_{F^0,j}\tilde{H}\check{G}\lambda_i\right)$$
$$+\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\tilde{H}^*\sum_{j\neq l}^{L}\overline{\overline{W}}_{F^0,j}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kl}\tilde{W}'_{F^0,l}\tilde{H}\check{G}\lambda_i\right)$$
$$+\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\hat{W}_F^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)$$
$$+\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)$$
$$\equiv \tilde{J}_{8,b1}+\tilde{J}_{8,b2}+\tilde{J}_{8,b3}+\tilde{J}_{8,b4}.$$

Next, define $\tilde{J}_{8,1}=\tilde{J}_{8,a1}+\tilde{J}_{8,b1}$, and define $\tilde{J}_{8,2}$ and $\tilde{J}_{8,3}$ analogously. Note that

$$\tilde{J}_{8,1} = -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\sum_{j=1}^{L}\overline{W}_{X,ij}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}'_{F^0,j}\tilde{H}\check{G}\lambda_i\right)$$
$$+\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(\frac{W_{X,i}^*\hat{W}_F}{T}\tilde{H}^*\sum_{j=1}^{L}\overline{\overline{W}}_{F^0,j}W_{\varepsilon,kj}\overline{W}_{\varepsilon,kj}\tilde{W}'_{F^0,j}\tilde{H}\check{G}\lambda_i\right)$$

Let $J_{it}=\left(X'_{it},F_t^{0\prime}\right)'$, 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, we can show 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 $\tilde{J}_{8,2}=O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$. In addition, we can show that

$$\tilde{J}_{8,1} = -\frac{1}{NT}\sum_{i=1}^{N}\text{Re}\left(W_{X,i}^*\left(\mathbb{I}_L-\hat{W}_F\left(\tilde{W}_{F^0}\tilde{H}\right)^*\right)\frac{1}{NT}\sum_{k=1}^{N}diag\left(|W_{\varepsilon,kj}|^2\right)\tilde{W}_{F^0}\tilde{H}\check{G}\lambda_i\right)$$
$$= -\frac{1}{NT}\sum_{i=1}^{N}\text{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}diag\left(|W_{\varepsilon,kj}|^2\right)\hat{W}_F\check{G}\lambda_i\right)+O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right)$$
$$= A_{NT}+O_p\left(\frac{1}{T}\gamma_L^{2+2d_{F,\min}-d_{X,\max}-3d_{F,\max}-2d_\varepsilon}\right),$$

by how we bound $\tilde{J}_{8,2}$. For $\tilde{J}_{8,3}$, we have

$$\tilde{J}_{8,3} = -\frac{1}{N^2T^2}\sum_{i,k=1}^{N}\text{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}W_{\varepsilon,k}W_{\varepsilon,k}^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)$$
$$= -\frac{1}{NT}\sum_{i=1}^{N}\text{Re}\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}\mathbb{E}\left(W_{\varepsilon,k}W_{\varepsilon,k}^*\right)\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)$$

$$
-\frac{1}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}\frac{1}{NT}\sum_{k=1}^{N}\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}-\mathbb{E}\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}\right)\right)\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\check{G}\lambda_i\right)
$$

$$
\equiv \tilde{J}_{8,31}+\tilde{J}_{8,32}.
$$

For $\tilde{J}_{8,31}$, we have

$$
\begin{aligned}
\left\|\tilde{J}_{8,31}\right\| &\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\frac{1}{NT}\sum_{i=1}^{N}\left\|\frac{W_{X,i}}{\sqrt{T}}\right\|\left\|\lambda_i\right\|\frac{1}{N}\sum_{k=1}^{N}\left\|\Omega_k\right\|_{sp}\left\|\check{\Gamma}_{\varepsilon}^{-1}\right\|_{sp}^{2}\frac{1}{\sqrt{T}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\| \\
&= O_p\left(T^{2d_{\varepsilon}-1}\gamma_L^{d_{F,\min}-d_{X,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_{\varepsilon}}\right)\right)
\end{aligned}
$$

by Assumption C*(iii). For $\tilde{J}_{8,32}$, we have

$$
\begin{aligned}
\left\|\tilde{J}_{8,32}\right\| &\lesssim \gamma_L^{d_{F,\min}-\frac{1}{2}}\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\left\|W_{X,i}\right\|\left\|\lambda_i\right\|\frac{1}{NT}\left\|\sum_{k=1}^{N}\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}-\mathbb{E}\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}\right)\right)\right\|\frac{1}{\sqrt{T}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\| \\
&= O_p\left(\gamma_L^{d_{F,\min}-\frac{1}{2}}\right)O_p\left(\gamma_L^{\frac{1}{2}-d_{X,\max}}\right)O_p\left(N^{-\frac{1}{2}}\gamma_L^{1-2d_{\varepsilon}}\right)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_{\varepsilon}}\right) \\
&= O_p\left(N^{-\frac{1}{2}}\gamma_L^{1-2d_{\varepsilon}+\left(d_{F,\min}-d_{X,\max}\right)}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_{\varepsilon}}\right)\right),
\end{aligned}
$$

because

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{NT}\left\|\sum_{k=1}^{N}\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}-\mathbb{E}\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}\right)\right)\right\|\right] \\
&\le \frac{1}{NT}\left(E\sum_{i,k=1}^{N}\operatorname{tr}\left[\left(W_{\varepsilon,i}W_{\varepsilon,i}^{*}-\mathbb{E}\left(W_{\varepsilon,i}W_{\varepsilon,i}^{*}\right)\right)\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}-\mathbb{E}\left(W_{\varepsilon,k}W_{\varepsilon,k}^{*}\right)\right)^{*}\right]\right)^{\frac{1}{2}} \\
&= \frac{1}{NT}\left(\sum_{i,k=1}^{N}\sum_{l,m=1}^{L}\operatorname{cov}\left(W_{\varepsilon,il}W_{\varepsilon,im}^{*},W_{\varepsilon,kl}W_{\varepsilon,km}^{*}\right)\right)^{\frac{1}{2}}=O\left(N^{-\frac{1}{2}}\gamma_L^{1-2d_{\varepsilon}}\right)
\end{aligned}
$$

by Assumption C*(v). In addition, $\tilde{J}_{8,b4}$ is dominated by $\tilde{J}_{8,b3}$ in order and is of asymptotically smaller order. This completes the proof of Lemma A.8. ■

Proof of Lemma A.9. We first show

$$
\frac{\sqrt{NL}\gamma_L^{d_{\varepsilon}-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^{*}\mathbf{M}_{\hat{W}_F}W_{\varepsilon,i}\right)=\frac{\sqrt{NL}\gamma_L^{d_{\varepsilon}-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left(W_{X,i}^{*}\mathbf{M}_{W_{F^0}}W_{\varepsilon,i}\right)+o_p(1).
$$

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

$$
\frac{\sqrt{NL}\gamma_L^{d_{\varepsilon}-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\operatorname{Re}\left[W_{X,i}^{*}\left(\mathbf{P}_{\hat{W}_F}-\mathbf{P}_{W_{F^0}}\right)W_{\varepsilon,i}\right]
$$

$$
\begin{aligned}
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\frac{\hat{W}_F\hat{W}_F^*}{T}W_{\varepsilon,i}-W_{X,i}^*\mathbf{P}_{\tilde{W}_{F^0}}W_{\varepsilon,i}\right)\\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\frac{\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^*\tilde{W}_{F^0}^*}{T}W_{\varepsilon,i}\right)\\
&+ \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\frac{\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*}{T}W_{\varepsilon,i}\right)\\
&+ \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\frac{\tilde{W}_{F^0}\tilde{H}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*}{T}W_{\varepsilon,i}\right)\\
&+ \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\left(\tilde{H}\tilde{H}^*-\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\right)\tilde{W}_{F^0}^*W_{\varepsilon,i}\right)\\
&\equiv a+b+c+d.
\end{aligned}
$$

We study a, b, c and d in turn. First, for a we have

$$
\begin{aligned}
\|a\| &= \left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\frac{\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\tilde{H}^*\tilde{W}_{F^0}^*}{T}W_{\varepsilon,i}\right)\right\|\\
&\leq \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|}{NT^{\frac{3}{2}}}\sum_{i=1}^{N}\|W_{X,i}\|\left\|\tilde{W}_{F^0}^*W_{\varepsilon,i}\right\|\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\frac{1}{T^{\frac{1}{2}}}\left\|\tilde{H}\right\|\\
&\lesssim \gamma_L^{1-2d_{F,\max}}\sqrt{NL}\gamma_L^{d_\varepsilon}\gamma_L^{-1}\|\Gamma_Z\|\left\{\frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^{N}\|W_{X,i}\|\left\|\tilde{W}_{F^0}^*W_{\varepsilon,i}\right\|\right\}\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\\
&= \gamma_L^{1-2d_{F,\max}}\sqrt{NL}\gamma_L^{d_\varepsilon}\gamma_L^{d_{Z,\min}-1}O_p\left(L^{-\frac{1}{2}}\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-2d_\varepsilon}\right)\\
&= \sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}O_p\left(\gamma_L^{2d_{Z,\min}-2d_{F,\max}+1-d_{X,\max}-d_\varepsilon}\delta_{W1,NT}\left\|\tilde{\delta}\right\|\right)\\
&+ O_p\left(\gamma_L^{2-2d_{F,\max}-2d_\varepsilon+d_{Z,\min}-d_{X,\max}}\right)\\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|\left\|\tilde{\delta}\right\|\right)+o_p(1),
\end{aligned}
$$

by Lemma A.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 L/T converges to zero sufficiently fast to represent an undersmoothed estimator.

Next, for b we have

$$
\|b\|=\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT^2}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,i}\right)\right\|
$$

$$\leq \sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\left\{\frac{1}{NT}\sum_{i=1}^{N}\left\|W_{X,i}\right\|\left\|W_{\varepsilon,i}\right\|\right\}\frac{1}{T}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|^2$$
$$=\sqrt{NL}\gamma_L^{d_\varepsilon-1}\gamma_L^{d_{Z,\min}}O_p\left(\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right)O_p\left(\delta_{W1,NT}^2\left\|\tilde{\delta}\right\|^2+N^{-1}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon}\right)$$
$$=\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}-d_\varepsilon}\delta_{W1,NT}^2\left\|\tilde{\delta}\right\|^2\right)+O_p\left(\sqrt{\frac{L}{N}}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon+d_{Z,\min}-d_{X,\max}}\right)$$
$$=o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|\left\|\tilde{\delta}\right\|\right)+o_p(1).$$

Next, for c we have

$$c=\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(W_{X,i}^*\frac{\tilde{W}_{F^0}\tilde{H}\tilde{H}^*\left(\hat{W}_F\tilde{H}^{-1}-\tilde{W}_{F^0}\right)^*}{T}W_{\varepsilon,i}\right)$$
$$=\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\left(\hat{W}_F\tilde{H}^{-1}-\tilde{W}_{F^0}\right)^*W_{\varepsilon,i}\right)$$
$$+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\tilde{Q}\left(\hat{W}_F\tilde{H}^{-1}-\tilde{W}_{F^0}\right)^*W_{\varepsilon,i}\right)\equiv c_1+c_2.$$

where $\tilde{Q}=\tilde{H}\tilde{H}^*-\left(\frac{1}{T}\tilde{W}_{F^0}^*\tilde{W}_{F^0}\right)^{-1}$. For c_2, we have

$$\left\|c_2\right\|=\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\tilde{Q}\tilde{H}^{*-1}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*W_{\varepsilon,i}\right\|$$
$$\leq\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\left\{\frac{1}{NT}\sum_{i=1}^{N}\left\|W_{X,i}\right\|\left\|W_{\varepsilon,i}\right\|\right\}\left\{\frac{1}{T^{\frac{1}{2}}}\left\|\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right\|\right\}\frac{1}{T^{\frac{1}{2}}}\left\|\tilde{W}_{F^0}\right\|\left\|\tilde{Q}\right\|\left\|\tilde{H}^{-1}\right\|$$
$$\leq O_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\gamma_L^{1-d_{X,\max}-d_\varepsilon}\right)O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\left\|\tilde{Q}\right\|\left\|\tilde{H}^{-1}\right\|.$$

Then it remains to study the order of $\left\|\tilde{Q}\right\|$. To do that, we consider

$$\frac{1}{T}\tilde{W}_{F^0}^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)=O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right),\tag{B.8}$$

and similarly

$$\frac{1}{T}\hat{W}_F^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)=O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right).\tag{B.9}$$

Then

$$\left\|\mathbb{I}_R-\tilde{H}^*\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\tilde{H}\right\|=\left\|\frac{1}{T}\hat{W}_F^*\hat{W}_{F^0}-\tilde{H}^*\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\tilde{H}\right\|\tag{B.10}$$
$$\leq\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\right\|+\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\tilde{W}_{F^0}\tilde{H}\right\|$$
$$+\left\|\tilde{H}^*\tilde{W}_{F^0}^*\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)\right\|$$

$$= \left\|\tilde{H}\right\| O_p\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right) = O_p\left(\gamma_L^{1-2d_{F,\max}}\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\| + N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\right).$$

and so is the probability order of $\left\|\tilde{Q}\right\|$. Therefore we conclude that

$$\begin{aligned}
c_2 &= \sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}O_p\left(\gamma_L^{2d_{Z,\min}-d_{X,\max}-d_\varepsilon+2d_{F,\min}-2d_{F,\max}}\delta_{W1,NT}^2\left\|\tilde{\delta}\right\|^2\right)\\
&\quad + O_p\left(\gamma_L^{d_{Z,\min}-d_{X,\max}+2d_{F,\min}-2d_{F,\max}}\sqrt{\frac{L}{N}}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon}\right)\\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|\left\|\tilde{\delta}\right\|\right) + o_p(1).
\end{aligned}$$

Next for c_1 we have

$$\begin{aligned}
c_1 &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N \mathrm{Re}\left(\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\left(\hat{W}_F\tilde{H}^{-1} - \tilde{W}_{F^0}\right)^* W_{\varepsilon,i}\right)\\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N \mathrm{Re}\left(\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\left(\frac{\tilde{W}_{F^0}^*\tilde{W}_{F^0}}{T}\right)^{-1}\tilde{\Gamma}_F\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\tilde{\Gamma}_F\left(\frac{\hat{W}_F^*\tilde{W}_{F^0}}{T}\right)^{-1}\left(\sum_{\ell=1}^8 \tilde{I}_\ell\right)^* W_{\varepsilon,i}\right)\\
&\equiv \sum_{\ell=1}^8 c_{1,\ell}.
\end{aligned}$$

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

$$\|c_{1,2}\| = \left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^N \frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}I_2^* W_{\varepsilon,i}\right\|$$

$$
\begin{aligned}
&= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} \hat{W}_F^* W_{F^0} \frac{1}{NT} \sum_{k=1}^{N} \lambda_k \tilde{\delta}' W_{X,k} W_{\varepsilon,i} \right\| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \Gamma_Z \right\| \left\{ \frac{1}{NT} \sum_{i=1}^{N} \left\| W_{X,i} \right\| \left\| W_{\varepsilon,i} \right\| \right\} \left\{ \frac{1}{NT^{\frac{1}{2}}} \sum_{k=1}^{N} \left\| W_{X,k} \right\| \left\| \lambda_k \right\| \right\} \left\{ \frac{1}{T^{\frac{1}{2}}} \left\| W_{F^0} \right\| \right\} \left\| \breve{G} \right\| \left\| \tilde{\delta} \right\| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon + d_{Z,\min} - 1} O_p\left(\gamma_L^{1 - d_{X,\max} - d_\varepsilon}\right) O_p\left(\gamma_L^{1/2 - d_{X,\max}}\right) O_p\left(\gamma_L^{\frac{1}{2} - d_{F,\max}}\right) O_p\left(\gamma_L^{2d_{F,\min} - 1}\right) \left\| \tilde{\delta} \right\| \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} O_p\left(\gamma_L^{2d_{F,\min} + 2d_{Z,\min} - d_{F,\max} - 2d_{X,\max} - d_\varepsilon} \left\| \tilde{\delta} \right\|\right) \\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\|\right).
\end{aligned}
$$

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

$$
\begin{aligned}
\left\| c_{1,6} \right\| &= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} I_6^* W_{\varepsilon,i} \right\| \\
&= \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} \hat{W}_F^* \frac{1}{NT} \sum_{k=1}^{N} W_{\varepsilon,k} \lambda_k' W_{F^0}^* W_{\varepsilon,i} \right\| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \Gamma_Z \right\| \frac{1}{NT} \sum_{i=1}^{N} \left\| W_{X,i} \right\| \left\| W_{F^0}^* W_{\varepsilon,i} \right\| \left\| \frac{1}{NT^{\frac{3}{2}}} \sum_{k=1}^{N} \hat{W}_F^* W_{\varepsilon,k} \lambda_k' \right\| \left\| \breve{G} \right\| \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon + d_{Z,\min} - 1} O_p\left(L^{-\frac{1}{2}} \gamma_L^{1 - d_{X,\max} - d_\varepsilon}\right) \\
&\times O_p\left(\delta_{W,NL} \left\| \tilde{\delta} \right\| + N^{-\frac{1}{2}} L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2} - 2d_{F,\max} - d_\varepsilon}\right) O_p\left(\gamma_L^{d_{F,\min} - d_{F,\max}}\right) \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} O_p\left(L^{-\frac{1}{2}} \gamma_L^{2d_{Z,\min} - d_{X,\max} - d_\varepsilon + d_{F,\min} - d_{F,\max}} \delta_{W,NL} \left\| \tilde{\delta} \right\|\right) \\
&+ O_p\left(L^{-\frac{1}{2}} \gamma_L^{\frac{3}{2} - 2d_{F,\max} - d_\varepsilon + d_{Z,\min} - d_{X,\max} + d_{F,\min} - d_{F,\max}}\right) \\
&= o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\|\right) + o_p(1)
\end{aligned}
$$

by Lemma A.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 non-random P-vector ω such that $\|\omega\| = 1$, we have

$$
\begin{aligned}
&\left| \omega' c_{1,7} \right| \\
&= \left| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon - 1} \omega' \Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^* \tilde{W}_{F^0}}{T} \breve{G} I_7^* W_{\varepsilon,i} \right| \\
&= \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \frac{1}{NT} \left| \operatorname{tr}\left(\tilde{W}_{F^0} \left(\frac{\tilde{W}_{F^0}^* \tilde{W}_{F^0}}{T} \right)^{-1} \tilde{\Gamma}_F \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \lambda_k W_{\varepsilon,k}^* \sum_{i=1}^{N} \frac{W_{\varepsilon,i} \omega' \Gamma_Z W_{X,i}^*}{T} \right) \right| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \tilde{\Gamma}_F \right\| \frac{1}{NT} \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k W_{\varepsilon,k}^* \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{W_{\varepsilon,i} \omega' \Gamma_Z W_{X,i}^* \tilde{W}_{F^0}}{T} \right\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\tilde{\Gamma}_F\right\|}{NT}\sum_{l=1}^{L}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}A_iW_{\varepsilon,il}\right\|\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_kW_{\varepsilon,kl}^*\right\| \\
&\leq \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\tilde{\Gamma}_F\right\|}{N}\left(\frac{1}{T}\sum_{l=1}^{L}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}A_iW_{\varepsilon,il}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{l=1}^{L}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_kW_{\varepsilon,kl}^*\right\|^2\right)^{\frac{1}{2}} \\
&= O_p\left(\sqrt{\frac{L}{N}}\gamma_L^{\frac{1}{2}+d_{Z,\min}+d_{F,\min}-d_{X,\max}-2d_\varepsilon}\right) = o_p(1),
\end{aligned}
$$

where $A_i = \frac{1}{T}\omega'\Gamma_Z W_{X,i}^*\tilde{W}_{F^0}$. 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:

$$
\begin{aligned}
\mathbb{E}\left(\frac{1}{T}\sum_{l=1}^{L}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}A_iW_{\varepsilon,il}\right\|^2\right) &\leq \frac{1}{NT}\sum_{l=1}^{L}\sum_{i,k=1}^{N}\left|\mathbb{E}\left(A_iA_k^*\right)\right|\left|\mathbb{E}\left(W_{\varepsilon,il}W_{\varepsilon,kl}^*\right)\right|, \\
&\leq \frac{1}{NT}\sum_{l=1}^{L}\sum_{i,k=1}^{N}\mathbb{E}\left|A_iA_k^*\right|\left(\mathbb{E}\left|W_{\varepsilon,il}W_{\varepsilon,kl}^*\right|^2\right)^{\frac{1}{2}} \qquad \text{(B.11)}
\end{aligned}
$$

where $\mathbb{E}\left|W_{\varepsilon,il}W_{\varepsilon,kl}^*\right|^2 \leq \gamma_l^{-4d_\varepsilon}\left(\overline{\sigma}_{ik}^W\right)^2$ by Assumption C*(ii). Note that $\mathbb{E}\left|A_iA_k^*\right| \leq \mathbb{E}\left(\|A_i\|\,\|A_k\|\right) \leq \sqrt{\mathbb{E}\|A_i\|^2\,\mathbb{E}\|A_k\|^2}$. Denote $\overline{W}_{X,ij}$ as the conjugate of $W_{X,ij}$, and $\check{W}_{F^0,j} = \check{\Gamma}_{F,j}\check{W}_{F^0,j}$ as in Assumption C*(iii). Define $\check{W}_{X,ij}$ analogously. Then

$$
\begin{aligned}
\mathbb{E}\|A_i\|^2 &= \frac{1}{T^2}\omega'\Gamma_Z\sum_{j,l=1}^{L}\mathbb{E}\left(\overline{W}_{X,ij}W_{F^0,j}'\tilde{\Gamma}_F^2W_{F^0,l}W_{X,il}^*\right)\Gamma_Z\omega \\
&= \frac{1}{T^2}\omega'\Gamma_Z\sum_{j,l=1}^{L}\mathbb{E}\left(\check{\Gamma}_{X,j}^{-1}\check{\overline{W}}_{X,ij}\check{W}_{F^0,j}'\check{\Gamma}_{F,j}^{-1}\tilde{\Gamma}_F^2\check{\Gamma}_{F,l}^{-1}\check{W}_{F^0,l}\check{W}_{X,il}^*\check{\Gamma}_{X,l}^{-1}\right)\Gamma_Z\omega \\
&\lesssim \frac{1}{T^2}\gamma_L^{2d_{Z,\min}}\sum_{j,l=1}^{L}\gamma_j^{-d_{X,\max}}\gamma_l^{-d_{X,\max}}\left\|\check{\Gamma}_{F,j}^{-1}\tilde{\Gamma}_F^2\check{\Gamma}_{F,l}^{-1}\right\|\mathbb{E}\left(\left\|\check{\overline{W}}_{X,ij}\check{W}_{F^0,j}'\right\|\left\|\check{W}_{F^0,l}\check{W}_{X,il}^*\right\|\right) \\
&= O\left(\gamma_L^{2-2d_{X,\max}+2d_{Z,\min}}\right)
\end{aligned}
$$

by Assumption B*(i) and B*(iii). In addition, $\frac{1}{NT}\sum_{l=1}^{L}\sum_{i,k=1}^{N}\left(\mathbb{E}\left|W_{\varepsilon,il}W_{\varepsilon,kl}^*\right|^2\right)^{\frac{1}{2}} = O\left(\gamma_L^{1-2d_\varepsilon}\right)$. So (B.11) is $O\left(\gamma_L^{3-2d_{X,\max}-2d_\varepsilon+2d_{Z,\min}}\right)$. Similarly, we have

$$
\mathbb{E}\left(\frac{1}{T}\sum_{l=1}^{L}\left\|\frac{1}{\sqrt{N}}\sum_{i=k}^{N}\lambda_kW_{\varepsilon,kl}\right\|^2\right) = O\left(\gamma_L^{1-2d_\varepsilon}\right),
$$

which altogether forms the order of $c_{1,7}$. Then lastly $c_{1,8}$ is given by

$$
\begin{aligned}
c_{1,8} &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}I_8^*W_{\varepsilon,i} \\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\hat{W}_F^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W_{\varepsilon,k}^*W_{\varepsilon,i} \\
&= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\tilde{H}^*\tilde{W}_{F^0}^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W_{\varepsilon,k}^*W_{\varepsilon,i} \\
&+ \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}W_{\varepsilon,k}^*W_{\varepsilon,i} \equiv c_{1,81}+c_{1,82}.
\end{aligned}
$$

Then it remains to study $c_{1,81}$ and $c_{1,82}$. For $c_{1,81}$, we have

$$
\begin{aligned}
c_{1,81} &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\tilde{H}^*\tilde{W}_{F^0}^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}\left[W_{\varepsilon,k}^*W_{\varepsilon,i}-\mathbb{E}\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right)\right] \\
&+ \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\breve{G}\tilde{H}^*\tilde{W}_{F^0}^*\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}\mathbb{E}\left(W_{\varepsilon,k}^*W_{\varepsilon,i}\right) \equiv c_{1,811}+c_{1,812}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\|c_{1,811}\| \\
&= \left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{k=1}^{N}\left[\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\sum_{l=1}^{L}\left[W_{\varepsilon,kl}^*W_{\varepsilon,il}-\mathbb{E}\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)\right]\right]\breve{G}\tilde{H}^*\left[\frac{1}{\sqrt{T}}\tilde{W}_{F^0}^*W_{\varepsilon,k}\right]\right\| \\
&\lesssim \sqrt{NL}\gamma_L^{d_\varepsilon-1}\|\Gamma_Z\|\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\sum_{l=1}^{L}\left[W_{\varepsilon,kl}^*W_{\varepsilon,il}-\mathbb{E}\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)\right]\right\|^2\right)^{\frac{1}{2}} \\
&\times\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\tilde{W}_{F^0}^*W_{\varepsilon,k}\right\|^2\right)^{\frac{1}{2}} \\
&= O_p\left(\sqrt{NL}\gamma_L^{d_{Z,\min}+d_\varepsilon-1}\right)O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_\varepsilon-d_{X,\max}}\right)O_p\left(T^{-\frac{1}{2}}L^{-\frac{1}{2}}\gamma_L^{\frac{1}{2}-d_\varepsilon}\right) \\
&= O_p\left(\frac{1}{\sqrt{T}}\gamma_L^{1-2d_\varepsilon+d_{Z,\min}-d_{X,\max}}\right) = o_p(1)
\end{aligned}
$$

by Assumption D*(iii), where the last two equalities hold by Assumption C*(i) and the fact that

$$
\mathbb{E}\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\frac{W_{X,i}^*\tilde{W}_{F^0}}{T}\sum_{l=1}^{L}\left[W_{\varepsilon,kl}^*W_{\varepsilon,il}-\mathbb{E}\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)\right]\right\|^2\right) = O\left(N^{-1}\gamma_L^{2-4d_\varepsilon}\gamma_L^{1-2d_{X,\max}}\right) \tag{B.12}
$$

by Assumption C*(v) following the similar reasoning to (B.3) in the proof of Lemma A.4. Similarly $c_{1,812}$ has the same order, which is obtained by replacing $W_{\varepsilon,kl}^*W_{\varepsilon,il}-\mathbb{E}\left(W_{\varepsilon,kl}^*W_{\varepsilon,il}\right)$ on the left hand side of (B.12) by $\mathbb{E}\left(W_{\varepsilon,kl}^{*}W_{\varepsilon,il}\right)$ and using Assumption C*(ii). For $c_{1,82}$, we have

$$\begin{aligned}c_{1,82} &= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}\left[W_{\varepsilon,k}^{*}W_{\varepsilon,i}-\mathbb{E}\left(W_{\varepsilon,k}^{*}W_{\varepsilon,i}\right)\right]\\&+\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}\mathbb{E}\left(W_{\varepsilon,k}^{*}W_{\varepsilon,i}\right)\\&\equiv c_{1,821}+c_{1,822},\end{aligned}$$

By Assumption B*(iii) and B*(iv), we have

$$\begin{aligned}&\|c_{1,821}\|\\&=\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}\left[W_{\varepsilon,k}^{*}W_{\varepsilon,i}-\mathbb{E}\left(W_{\varepsilon,k}^{*}W_{\varepsilon,i}\right)\right]\right\|\\&\lesssim\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\gamma_L^{2d_{F,\min}-1}\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\sum_{l=1}^{L}\left[W_{\varepsilon,kl}^{*}W_{\varepsilon,il}-\mathbb{E}\left(W_{\varepsilon,kl}^{*}W_{\varepsilon,il}\right)\right]\right\|^2\right)^{\frac{1}{2}}\\&\times\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}W_{\varepsilon,k}\right\|^2\right)^{\frac{1}{2}}\\&=\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|O_p\left(\gamma_L^{2d_{F,\min}-1}\right)O_p\left(N^{-\frac{1}{2}}\gamma_L^{\frac{3}{2}-2d_\varepsilon-d_{X,\max}}\right)O_p\left(\left(\delta_{W1,NT}\left\|\tilde{\delta}\right\|+N^{-\frac{1}{2}}\gamma_L^{1-d_{F,\max}-d_\varepsilon}\right)\gamma_L^{\frac{1}{2}-d_\varepsilon}\right)\\&=O_p\left(\alpha_{3,NL}\left\|\tilde{\delta}\right\|+\sqrt{\frac{L}{N}}\gamma_L^{1-d_{X,\max}-d_{F,\max}+d_{Z,\min}+2d_{F,\min}-3d_\varepsilon}\right)=o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\left\|\Gamma_Z^{-1}\right\|\left\|\tilde{\delta}\right\|\right)+o_p(1)\end{aligned}$$

by Lemma A.6, as

$$\mathbb{E}\left(\frac{1}{NT}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}W_{\varepsilon,k}\right\|^2\right)=O\left(\left(\delta_{W1,NT}^2\left\|\tilde{\delta}\right\|^2+N^{-1}\gamma_L^{2-2d_{F,\max}-2d_\varepsilon}\right)\gamma_L^{1-2d_\varepsilon}\right),$$

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

$$\begin{aligned}&\|c_{1,822}\|\\&=\left\|\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\breve{G}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}\frac{1}{NT}\sum_{k=1}^{N}W_{\varepsilon,k}E\left(W_{\varepsilon,k}^{*}W_{\varepsilon,i}\right)\right\|\\&\lesssim\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\gamma_L^{2d_{F,\min}-1}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\right\|\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}W_{\varepsilon,k}\right\|\left|\frac{E\left(W_{\varepsilon,k}^{*}W_{\varepsilon,i}\right)}{T}\right|\\&\leq\sqrt{NL}\gamma_L^{d_\varepsilon-1}\left\|\Gamma_Z\right\|\gamma_L^{2d_{F,\min}-1}\left\{\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{T}\left(\hat{W}_F-\tilde{W}_{F^0}\tilde{H}\right)^{*}W_{\varepsilon,k}\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{W_{X,i}^{*}\tilde{W}_{F^0}}{T}\right\|^2\right\}^{1/2}\end{aligned}$$

$$\times \left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{k=1}^{N} \left\| \frac{E\left(W_{\varepsilon,k}^{*} W_{\varepsilon,i} \right)}{T} \right\|^{2} \right\}^{1/2}$$

$$= \sqrt{NL}\gamma_L^{d_\varepsilon - 1} \left\| \Gamma_Z \right\| \gamma_L^{2d_{F,\min}-1} O_p \left(\left(\delta_{W1,NT} \left\| \tilde{\delta} \right\| + N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon} \right) \gamma_L^{\frac{1}{2}-d_\varepsilon} \right) O_p \left(\gamma_L^{\frac{1}{2}-d_{X,\max}} \right) O_p \left(N^{-\frac{1}{2}} \gamma_L^{1-2d_\varepsilon} \right)$$

$$= \sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} O_p \left(N^{-\frac{1}{2}} \gamma_L^{2d_{Z,\min}+2d_{F,\min}-d_{X,\max}-3d_\varepsilon} \delta_{W1,NT} \left\| \tilde{\delta} \right\| \right) + O_p \left(\sqrt{\frac{L}{N}} \gamma_L^{1+d_{Z,\min}+2d_{F,\min}-d_{X,\max}-d_{F,\max}-3d_\varepsilon} \right)$$

$$= o_p \left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right) + o_p (1)$$

by Lemma A.6 and by Assumption C*(ii).

Lastly for d we have

$$\|d\| = \left\| \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \frac{W_{X,i}^{*} \tilde{W}_{F^0}}{T} \tilde{Q} \tilde{W}_{F^0}^{*} W_{\varepsilon,i} \right\|$$

$$\leq \sqrt{NL}\gamma_L^{d_\varepsilon-1} \left\| \Gamma_Z \right\| \left\| \tilde{Q} \right\| \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{W_{X,i}^{*} \tilde{W}_{F^0}}{T} \right\| \left\| \frac{\tilde{W}_{F^0}^{*} W_{\varepsilon,i}}{T} \right\|$$

$$\leq \sqrt{NL}\gamma_L^{d_\varepsilon-1} \left\| \Gamma_Z \right\| \left\| \tilde{Q} \right\| \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{W_{X,i}^{*} \tilde{W}_{F^0}}{T} \right\|^{2} \right\}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\tilde{W}_{F^0}^{*} W_{\varepsilon,i}}{T} \right\|^{2} \right\}^{1/2}$$

$$= \sqrt{NL}\gamma_L^{d_\varepsilon-1} \left\| \Gamma_Z \right\| O_p \left(\gamma_L^{1-2d_{F,\max}} \left(\delta_{W1,NT} \left\| \tilde{\delta} \right\| + N^{-\frac{1}{2}} \gamma_L^{1-d_{F,\max}-d_\varepsilon} \right) \right)$$

$$\times O_p \left(\gamma_L^{\frac{1}{2}-d_{X,\max}} \right) O_p \left(L^{-\frac{1}{2}} \gamma_L^{\frac{1}{2}-d_\varepsilon} \right)$$

$$= \sqrt{NL}\gamma_L^{d_\varepsilon} \Gamma_Z^{-1} O_p \left(L^{-\frac{1}{2}} \gamma_L^{1-d_{X,\max}-d_\varepsilon+2d_{Z,\min}-2d_{F,\max}} \delta_{W1,NT} \left\| \tilde{\delta} \right\| \right) + O_p \left(\gamma_L^{2-d_{X,\max}-d_{F,\max}-2d_\varepsilon+d_{Z,\min}-d_{F,\max}} \right)$$

$$= o_p \left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right) + o_p (1),$$

by Assumption D*(iii) as before.

This completes the proof of approximation for the first part as

$$\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re} \left(W_{X,i}^{*} \mathbf{M}_{\hat{W}_F} W_{\varepsilon,i} \right)$$

$$= \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re} \left(W_{X,i}^{*} \mathbf{M}_{W_{F^0}} W_{\varepsilon,i} \right) + o_p \left(\sqrt{NL}\gamma_L^{d_\varepsilon} \left\| \Gamma_Z^{-1} \right\| \left\| \tilde{\delta} \right\| \right) + o_p (1).$$

The second part is given by

$$\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re} \left[\left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} W_{X,k}^{*} \right) \left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{W_{F^0}} \right) W_{\varepsilon,i} \right]$$

$$\equiv \frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT} \sum_{i=1}^{N} \operatorname{Re} \left[W_{V,i}^{*} \left(\mathbf{M}_{\hat{W}_F} - \mathbf{M}_{W_{F^0}} \right) W_{\varepsilon,i} \right].$$

By replacing $W_{X,i}$ by $W_{V,i}$, we can obtain the same order for the second part. Then we conclude that

$$\begin{aligned}
&\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\left(W_{X,i}^*\mathbf{M}_{\hat{W}_F}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}W_{X,k}^*\mathbf{M}_{\hat{W}_F}\right)W_{\varepsilon,i}\right)\\
&=\frac{\sqrt{NL}\gamma_L^{d_\varepsilon-1}\Gamma_Z}{NT}\sum_{i=1}^{N}\mathrm{Re}\left(\left(W_{X,i}^*\mathbf{M}_{W_{F^0}}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}W_{X,k}^*\mathbf{M}_{W_{F^0}}\right)W_{\varepsilon,i}\right)\\
&+o_p\left(\sqrt{NL}\gamma_L^{d_\varepsilon}\Gamma_Z^{-1}\left\|\tilde{\delta}\right\|\right)+o_p(1).
\end{aligned}$$

This completes the proof of Lemma A.9. ■

C Demeaned Time Domain Least Squares Estimation

In this section, we briefly study an alternative estimation method in time domain, the LS estimation based on the within-group demeaned equation. Such analysis can help understand in another perspective the complexity in time domain LS estimation due to long memory.

By (2.2) and (2.3), the model in (2.1) can be rewritten as

$$\begin{aligned}
Y_{it}&=X_{it}'\beta^0+\lambda_i'F_t^0+\varepsilon_{it}\\
&=\left(\mu_{X,i}+X_{it}^o\right)'\beta^0+\lambda_i'\left(\mu_F+F_t^o\right)+\varepsilon_{it}\\
&=X_{it}^{o\prime}\beta^0+\lambda_i'F_t^0+\tilde{\mu}_i+\varepsilon_{it},
\end{aligned}\tag{C.1}$$

where $\tilde{\mu}_i=\mu_{X,i}'\beta^0+\lambda_i'\mu_F$ is an additive individual effect. Then following Bai (2009), we conduct the LS estimation to its demeaned (with-group transformed) version

$$\dot{Y}_{it}=\dot{X}_{it}^{o\prime}\beta^0+\lambda_i'\dot{F}_t^o+\dot{\varepsilon}_{it},\tag{C.2}$$

where $\dot{Y}_{it}=Y_{it}-\overline{Y}_{i\cdot}$, $\overline{Y}_{i\cdot}=\frac{1}{T}\sum_{t=1}^{T}Y_{it}$, and $\dot{X}_{it}^o$, $\dot{F}_t^o$, and $\dot{\varepsilon}_{it}$ are analogously defined. Let $\dot{X}_i^o=(\dot{X}_{i1}^o,...\dot{X}_{iT}^o)'$ and $\dot{F}^o=(\dot{F}_1^o,...,\dot{F}_T^o)'$. Define $\dot{Y}_i$ analogously. We consider the PCA based on (C.2). The LS estimators β^* and F^* of β^0 and $\dot{F}^o$ solve the system of nonlinear equations:

$$\beta^*=\left(\sum_{i=1}^{N}\dot{X}_i^{o\prime}\mathbf{M}_{F^*}\dot{X}_i^o\right)^{-1}\sum_{i=1}^{N}\dot{X}_i^{o\prime}\mathbf{M}_{F^*}\dot{Y}_i\tag{C.3}$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\dot{Y}_i-\dot{X}_i^o\beta^*\right)\left(\dot{Y}_i-\dot{X}_i^o\beta^*\right)'\right]F^*=F^*V_{NT}.\tag{C.4}$$

Let $Z_i^*=\mathbf{M}_{\dot{F}^o}\dot{X}_i^o-\frac{1}{N}\sum_{k=1}^{N}a_{ik}\mathbf{M}_{\dot{F}^o}\dot{X}_k^o$.

The following theorem presents the asymptotic distribution of β^*.

Theorem C.1 *Suppose that Assumptions A-F hold and $T/N \to \rho \in (0,\infty)$ as $(N,T)\to\infty$. Let C^* denote the probability limit of*

$$\tilde{C}^* = -D\left(\dot{F}^o\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\dot{X}_i^o-\dot{V}_i^o\right)'\dot{F}^o}{T}\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{T}\sum_{k=1}^{N}\lambda_k\dot{\varepsilon}_k'\dot{\varepsilon}_i$$

with $\dot{V}_i^o = \frac{1}{N}\sum_{k=1}^{N}a_{ik}\dot{X}_k^o$. Then for some positive definite matrices $\tilde{D}_0$ and $\tilde{\Sigma}$ we have

(i) when $d_{Z,\max}+d_\varepsilon > \frac{1}{2}$ and $d_{F,\max}+d_\varepsilon > \frac{1}{2}$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^*-\beta^0-\frac{1}{N}C^*-\frac{1}{T^{2-d_{Z,\max}-d_{F,\max}-2d_\varepsilon}}A_1^*\right)\xrightarrow{d}\mathcal{N}\left(0,\tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where A_1^* is the probability limit of

$$\tilde{A}_1^* = -D\left(\dot{F}^o\right)^{-1}\frac{1}{NT^{d_{Z,\max}+d_\varepsilon}}\sum_{i=1}^{N}\dot{X}_i^{o\prime}\mathbf{M}_{\dot{F}^o}\frac{1}{NT^{d_{F,\max}+d_\varepsilon}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k'\dot{F}^o\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i;$$

(ii) when $d_{Z,\max}+d_\varepsilon > \frac{1}{2} \geq d_{F,\max}+d_\varepsilon$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^*-\beta^0-\frac{1}{N}C^*-\frac{1}{T^{1-\left(d_{Z,\max}+d_\varepsilon\right)}T^{\frac{1}{2}}}A_2^*\right)\xrightarrow{d}\mathcal{N}\left(0,\tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where A_2^* is the probability limit of

$$\tilde{A}_2^* = -D\left(\dot{F}^o\right)^{-1}\frac{1}{NT^{d_{Z,\max}+d_\varepsilon}}\sum_{i=1}^{N}\dot{X}_i^{o\prime}\mathbf{M}_{\dot{F}^o}\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k'\dot{F}^o\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i;$$

(iii) when $d_{F,\max}+d_\varepsilon > \frac{1}{2} \geq d_{Z,\max}+d_\varepsilon$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^*-\beta^0-\frac{1}{N}C^*-\frac{1}{T^{1-\left(d_{F,\max}+d_\varepsilon\right)}T^{\frac{1}{2}}}A_3^*\right)\xrightarrow{d}\mathcal{N}\left(0,\tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where A_3^* is the probability limit of

$$\tilde{A}_3^* = -D\left(\dot{F}^o\right)^{-1}\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\dot{X}_i^{o\prime}\mathbf{M}_{\dot{F}^o}\frac{1}{NT^{d_{F,\max}+d_\varepsilon}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k'\dot{F}^o\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i;$$

(iv) when $d_{Z,\max}+d_\varepsilon \leq \frac{1}{2}$ and $d_{F,\max}+d_\varepsilon \leq \frac{1}{2}$,

$$N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}\left(\beta^*-\beta^0-\frac{1}{N}C^*-\frac{1}{T}A_4^*\right)\xrightarrow{d}\mathcal{N}\left(0,\tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}\right),$$

where A_4^* is the probability limit of

$$\tilde{A}_4^* = -D\left(\dot{F}^o\right)^{-1}\frac{1}{NT^{\frac{1}{2}}}\sum_{i=1}^{N}\dot{X}_i^{o\prime}\mathbf{M}_{\dot{F}^o}\frac{1}{NT^{\frac{1}{2}}}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k'\dot{F}^o\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i.$$

With the within-group transformation, the LS estimator now obtains a unified convergence rate that only depends on the memory parameter of idiosyncratic error term. For each case, we have two bias terms where $\frac{1}{N}B^*$ is common and asymptotically negligible when $d_\varepsilon = 0$ and the order of the other bias term depends both on $d_{Z,\max} + d_\varepsilon$ and $d_{F,\max} + d_\varepsilon$. The latter bias feature makes it difficult to implement the LS estimator in the time domain.

Proof of Theorem C.1. The proof of Theorem C.1 follows the same steps used in the proof of Theorem 3.1. It is easy to see that all the asymptotically negligible terms there are still negligible here, and thus we can focus on the order of bias terms and the convergence rate of β^*. Specifically, the orders of two bias terms of the LS estimator of model (C.2), C^* and A_j^*, $j = 1, \ldots, 4$, are respectively related to the orders of the following two terms

$$\widetilde{c}_{1,7}^* = -\frac{1}{NT}\sum_{i=1}^{N}\frac{\left(\dot{X}_i^o - \dot{V}_i^o\right)'\dot{F}^o}{T}\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{N}\sum_{k=1}^{N}\lambda_k\dot{\varepsilon}_k'\dot{\varepsilon}_i, \tag{C.5}$$

and

$$\widetilde{J}_8^* = -\frac{1}{NT}\sum_{i=1}^{N}\dot{X}_i^{o\prime}\mathbf{M}_{\dot{F}^o}\frac{1}{NT}\sum_{k=1}^{N}\dot{\varepsilon}_k\dot{\varepsilon}_k'\dot{F}^o\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i. \tag{C.6}$$

First, for (C.5), we can use the same arguments as in the non-demeaned model to obtain $\widetilde{c}_{1,7}^* = O_p\left(\frac{1}{N}\right)$. For (C.6), we firstly denote $\dot{Z}_i^{o\prime} = \dot{X}_i^{o\prime}\mathbf{M}_{\dot{F}^o}$. Then by the definition of Z_i^*, we can see the memory parameter vector of $\dot{\tilde{Z}}_i$ is d_Z, which implies that

$$\widetilde{J}_8^* = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\dot{\tilde{Z}}_{it}\dot{\varepsilon}_{kt}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\dot{F}_t^o\dot{\varepsilon}_{kt}\right)\left(\frac{\dot{F}^{o\prime}\dot{F}^o}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,$$

where $\frac{1}{T}\sum_{t=1}^{T}\dot{Z}_{it}^o\dot{\varepsilon}_{kt}$ (c.f. $\frac{1}{T}\sum_{t=1}^{T}\dot{F}_t^o\dot{\varepsilon}_{kt}$) can be treated as the sample cross-covariance between $\dot{Z}_{it}^o$ (c.f. $\dot{F}_t^o$) 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 further implies that

$$\widetilde{J}_8^* = \begin{cases} O_p\left(T^{d_{Z,\max}+d_\varepsilon-1}T^{d_{F,\max}+d_\varepsilon-1}\right) \text{ if } d_{Z,\max}+d_\varepsilon > \frac{1}{2} \text{ and } d_{F,\max}+d_\varepsilon > \frac{1}{2} \\ O_p\left(T^{d_{Z,\max}+d_\varepsilon-1}T^{-\frac{1}{2}}\right) \text{ if } d_{Z,\max}+d_\varepsilon > \frac{1}{2} \ge d_{F,\max}+d_\varepsilon \\ O_p\left(T^{d_{F,\max}+d_\varepsilon-1}T^{-\frac{1}{2}}\right) \text{ if } d_{F,\max}+d_\varepsilon > \frac{1}{2} \ge d_{Z,\max}+d_\varepsilon \\ O_p(T^{-1}) \text{ if } d_{Z,\max}+d_\varepsilon \le \frac{1}{2} \text{ and } d_{F,\max}+d_\varepsilon \le \frac{1}{2} \end{cases}.$$

Then the asymptotic representation of $\beta^* - \beta^0$ follows that

$$\beta^* - \beta^0 = D\left(\dot{F}^o\right)^{-1}\left[\frac{1}{NT}\sum_{i=1}^{N}Z_i^{*\prime}\dot{\varepsilon}_i + \widetilde{c}_{1,7}^* + \widetilde{J}_8^*\right] + o_p\left(a_{NT}^{-1}\right),$$

and Assumption F implies that $N^{-\frac{1}{2}}T^{d_\varepsilon-\frac{1}{2}}\sum_{i=1}^{N}Z_i^{*\prime}\dot{\varepsilon}_i \xrightarrow{d} \mathcal{N}\left(0,\tilde{\Sigma}\right)$. In addition, $D\left(\dot{F}^o\right) \xrightarrow{p} \tilde{D}_0 > 0$. Then we have

$$\begin{cases} a_{NT}\left(\beta^* - \beta^0 - \frac{1}{N}C^* - \frac{1}{T^{2-(d_{F,\max}+d_{Z,\max}+2d_\varepsilon)}}A_1^*\right) \xrightarrow{d} \mathcal{N}(0,\tilde{\Omega}_0) \text{ if } d_{Z,\max}+d_\varepsilon > \frac{1}{2} \ \& \ d_{F,\max}+d_\varepsilon > \frac{1}{2} \\ a_{NT}\left(\beta^* - \beta^0 - \frac{1}{N}C^* - \frac{1}{T^{1-\left(d_{Z,\max}+d_\varepsilon\right)}T^{\frac{1}{2}}}A_2^*\right) \xrightarrow{d} \mathcal{N}(0,\tilde{\Omega}_0) \text{ if } d_{Z,\max}+d_\varepsilon > \frac{1}{2} \geq d_{F,\max}+d_\varepsilon \\ a_{NT}\left(\beta^* - \beta^0 - \frac{1}{N}C^* - \frac{1}{T^{1-\left(d_{F,\max}+d_\varepsilon\right)}T^{\frac{1}{2}}}A_3^*\right) \xrightarrow{d} \mathcal{N}(0,\tilde{\Omega}_0) \text{ if } d_{F,\max}+d_\varepsilon > \frac{1}{2} \geq d_{Z,\max}+d_\varepsilon \\ a_{NT}\left(\beta^* - \beta^0 - \frac{1}{N}C^* - \frac{1}{T}A_4^*\right) \xrightarrow{d} \mathcal{N}(0,\tilde{\Omega}_0) \text{ if } d_{Z,\max}+d_\varepsilon \leq \frac{1}{2} \ \& \ d_{F,\max}+d_\varepsilon \leq \frac{1}{2} \end{cases},$$

where $a_{NT} = N^{\frac{1}{2}}T^{\frac{1}{2}-d_\varepsilon}$, $\tilde{\Omega}_0 = \tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1}$, and C^*, A_1^*, A_2^*, A_3^* and A_4^* are as defined in the theorem. This completes the proof of Theorem C.1. ■