ISSN 1440-771X

Department of Econometrics and Business Statistics

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

Time Series Forecasting using a Mixture of Stationary and Nonstationary Predictors

Sium Bodha Hannadige, Jiti Gao, Mervyn J Silvapulle and Param Silvapulle

April 2021

Working Paper 06/21

(Revised version of working paper 19/20)

Time Series Forecasting using a Mixture of Stationary and Nonstationary Predictors

Sium Bodha Hannadige[*], Jiti Gao, Mervyn J Silvapulle, and

Param Silvapulle
Department of Econometrics and Business Statistics, Monash University, Australia [†]

Abstract

We develop a method for constructing prediction intervals for a nonstationary variable, such as GDP. The method uses a factor augmented regression [FAR] model. The predictors in the model includes a small number of factors generated to extract most of the information in a set of panel data on a large number of macroeconomic variables considered to be potential predictors. The novelty of this paper is that it provides a method and justification for a mixture of stationary and nonstationary factors as predictors in the FAR model; we refer to this as *mixture-FAR* method. This method is important because typically such a large set of panel data, for example the FRED-MD, is likely to contain a mixture of stationary and nonstationary variables. In our simulation study, we observed that the proposed mixture-FAR method performed better than its competitor that requires all the predictors to be nonstationary; the MSE of prediction was at least 33% lower for mixture-FAR. Using the data in FRED-QD for the US, we evaluated the aforementioned methods for forecasting the nonstationary variables, GDP and Industrial Production. We observed that the mixture-FAR method performed better than its competitors.

Keywords: Gross domestic product; high dimensional data; industrial production; macroeconomic forecasting; panel data.

JEL Classifications: C22, C33, C38, C53

[*]The authors gratefully acknowledge participants at several seminars and international workshops for their constructive comments and suggestions. This research was partially supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme under Grant Numbers: DP150101012 and DP170104421 for the second author, and DP150100210 for the third author).

[†]Email addresses for authors: sium.bodhahannadige@monash.edu, jiti.gao@monash.edu, mervyn.silvapulle@monash.edu, and param.silvapulle@monash.edu.

1 Introduction

Construction of valid probability forecasts of key economic variables, such as GDP and Inflation, is central to making reliable economic policy decisions. There is a large body of literature on constructing probability forecasts for a stationary variable using other stationary variables as predictors. By contrast, the literature on making probability forecasts for a nonstationary variable using a mixture of stationary and nonstationary predictors remains underdeveloped. In a method that has attracted considerable attention, a two-step method involving a *factor model* for panel data and a *regression model* for predicting the time series are used jointly (Stock & Watson 2002a). In the first step, the factor model is used for generating a small number of factors to capture most of the information in a set of panel data for a large number of potential predictors. In the second step, the regression model uses the generated factors as predictors, instead of the large number of potential predictors in the panel data. The resulting regression model is known as *factor augmented regression*[FAR] model, which is one of the well-known models for constructing probability forecasts for a time series (Bernanke et al. 2005, Stock & Watson 1998a, 1998b, 2002b). The large number of economic variables that are potential predictors typically includes a mixture of stationary and nonstationary variables. Consequently, the collection of factors is also typically a mixture of stationary and nonstationary ones (Bai 2004, Eickmeier 2005, Moon & Perron 2007, Smeekes & Wijler 2019). The objective of this paper is to develop a new method for constructing a valid prediction interval when the predictors in the prediction model include a mixture of stationary and nonstationary factors.

For the main results of this paper, the only nonstationary variables considered are $I(1)$; therefore, we use the term nonstationary as a synonym for $I(1)$.

Related literature

The validity of the aforementioned general approach for forecasting using an FAR model with estimated factors has been established when all the variables, including the factors, are stationary (Bai 2003, Bai & Ng [2002, 2006], Gonçalves & Perron 2014), and also when they are all nonstationary (Choi 2017), but not when they form a mixture of stationary and nonstationary ones. This paper builds on the aforementioned literature and develops a method based on FAR models for forecasting, more specifically for constructing an asymptotically valid prediction interval when the chosen set of factors is a mixture of stationary and nonstationary ones.

Suppose that the variables are all nonstationary. Bai (2004) studied the consistency of the estimated factors and proposed a method for estimating an optimal number of factors. The limiting distributions of the estimators of factors and their loadings have also been obtained. Choi (2017) used a method based on generalized principal components for estimating factors, and studied the asymptotic properties of the generated nonstationary factors, their loadings, and forecasts. Under the assumption $T/N \to 0$, Choi (2017) showed that estimators of the parameters in the forecasting model are consistent and asymptotically normal, and that the forecasts converge at the rate T, where T and N are the time and cross-section dimensions respectively.

Since the method in this paper is based on the large literature for forecasting a stationary variable using an FAR model, a few comments would be helpful. Suppose that we wish to predict a stationary variable, such as inflation, using a method that requires all the predictors in the prediction equation to be stationary, for example the method in Bai (2003) or that in Bai & Ng (2006). For this scenario, one could either delete all the I(1) variables or use the first differences of the I(1) predictors instead of the original I(1) predictors (Ludvigson & Ng 2007, Stock & Watson 2012, Cheng & Hansen 2015). While this adaptation is methodologically valid, a natural question that arises is whether differencing a nonstationary variable could result in loss of information in the level-data that may be important for forecasting. Similar questions also arise when forecasting a nonstationary variable, the topic of this paper.

Suppose that the set of generated factors is a mixture of stationary and nonstationary variables, and we wish to predict a nonstationary variable, such as GDP, using a method that requires all the predictors to be nonstationary, for example the method in Choi (2017). For this scenario, it has been suggested to delete all the predictors that are stationary and apply the method. While this method is valid, deletion of predictors to suit a method is likely to result in loss of information and hence loss of statistical efficiency.

The development of methodology for factor models has contributed to improve time series forecasting, macroeconomic analysis, and monetary policy analysis. Empirical results from several studies indicate that the generated factors often tend to be a mixture of stationary and nonstationary variables. For example, Bai (2004) studied employment fluctuations across 60 industries in the US and found that two nonstationary and one stationary factors explain a large part of the fluctuations in employment. Bernanke et al. (2005) used factor augmented vector auto-regression and found that it contained information to accurately identify the monetary transmission mechanism in the US. Eickmeier (2005) used a large-scale ($N > 300$) dynamic factor model and concluded that the Euro-area economies shared four non-stationary factors and one stationary factor. Eickmeier found that the factors represent mainly the variations in German and French real economic activity as well as of producer prices and financial prices through which they also studied the transmission channels and the impacts of macroeconomic shocks. Moon & Perron (2007) studied the Canadian and US interest rates for different maturities and risk, and found a single nonstationary factor and several stationary ones. The dominant factors were interpreted as level and slope, as in the term structure literature. In a recent study, Smeekes & Wijler (2019) provided an overview of forecasting macroeconomic time series in the presence of unit roots and cointegration. They compared point forecasts of some key economic variables in FRED-MD and FRED-QD data set and nowcasting of unemployment in another data set that was constructed from Google trend using the two methods (a) transforming every series to stationarity, and (b) directly modelling the level data. However, rigorous justification for modelling the level data with unit roots and cointegration in the forecasting model is yet to be provided.

The method in this paper

In this paper, we use the methods in the literature (Bai 2004, and Moon & Perron 2007) for generating factors that may be a mixture of stationary and nonstationary variables. Once they have been generated, we use them as predictors in a *factor-augmented regression* [FAR] model for forecasting; we refer to this as a *mixture-FAR* model. We develop new methods for constructing asymptotically valid prediction intervals using the mixture-FAR model. Our results, under the additional assumption that all the variables are stationary, reduce to the corresponding ones in Bai & Ng (2006). Similarly, our results, under the additional assumption that all the variables are nonstationary, reduce to the corresponding ones in Choi (2017). In this sense, our results provide a way of combining and extending the existing results on this topic that are limited to the two cases (a) when all the variables are stationary and (b) when all the variables are nonstationary.

To state the asymptotic results, we introduce a diagonal matrix, denoted D_{1T}; its dimension is equal to the number of predictors in the FAR model, and each of its diagonal element is equal to either $\sqrt{T}$ or T according as the corresponding predictor is stationary or nonstationary. The joint limiting distribution of the generated factors is derived under the assumption $\sqrt{N}\left\|D_{1T}^{-2}\right\| \to 0$, where and in what follows $\|A\| = trace(A'A)^{1/2}$. We develop the main part of the asymptotic results under the assumption $T/N \to 0$. We show the consistency and asymptotic normality of estimators of the parameters of the forecasting model. For the case of normally distributed errors in the prediction model with $\sqrt{N}\left\|D_{1T}^{-2}\right\| \to 0$ and $T/N \to 0$, we show that forecast error has an asymptotically normal distribution, and use it to construct an asymptotically valid prediction interval for the dependent variable in the forecast equation. To examine the finite sample properties of the estimates, we conducted a simulation study with data generating processes [DGP] that contain mixtures of stationary and nonstationary variables. In these simulations, we observed that the mixture-FAR method performed overall better than the method that requires all the variables to be nonstationary. As an empirical illustration, we evaluated the aforementioned methods for forecasting the nonstationary variables, GDP and industrial production [IP], using the quarterly panel data on US macroeconomic variables, known as FRED-QD. We observed that the mixture-FAR model performed better than its aforementioned competitors. This observation also corroborates the general observation of our simulation study, namely, the mixture-FAR method performed better than the competing methods.

The rest of this paper is organized as follows. Section 2 introduces the model and the assumptions, and establish the consistency and limiting distributions of the estimators. Section 3 reports the results of the simulation study. The empirical example using the FRED-QD data is presented in Section 4, and Section 5 concludes. The proofs of the theorems and lemmas, and some simulation results are provided in the Supplementary Materials to this paper available online.

2 Methodology

2.1 Model and notation

Let $\{Y_t, t = 1, 2, \ldots\}$ denote an observable univariate time series that we wish to predict at a future time $T+h$ $(h \geq 1)$, using the information available up to time T. Let $\{X_{it} \in \mathbb{R} : i = 1, \ldots, N; t = 1, \ldots, T\}$ denote a set of panel data and $\{W_t \in \mathbb{R}^m : t = 1, \ldots, T\}$ denote a set of observable predictors; W_t may contain lagged values of Y_t. The aforementioned *factor augmented regression*[FAR] method for predicting Y_{T+h} uses the following two models:

$$\text{Factor model:} \quad X_{it} = \lambda_i' F_t + e_{it} \qquad (i = 1, ..., N; t = 1, ..., T) \tag{1}$$

$$\text{FAR model:} \quad Y_{t+h} = \theta' F_t + \omega' W_t + \epsilon_{t+h} \qquad (t = 1, ..., T), \tag{2}$$

where F_t is an $r \times 1$ vector of unobservable factors, $\{e_{it}, \epsilon_t\}$ are idiosyncratic errors, λ_i is an $r \times 1$ vector of factor loadings, and $\theta_{r\times 1}$ and $\omega_{m\times 1}$ are unknown parameters $(i = 1, \ldots, N; t = 1, \ldots, T)$; the number of factors r is assumed known. This was called a "diffusion index forecasting model" by Stock & Watson (2002a).

A point of departure of our paper from the current literature is that we allow the r factors to be used in the FAR-method, to be a mixture of stationary and nonstationary variables. Further, we assume that Y_t and W_t are nonstationary; as indicated previously, the only nonstationary variables that we consider are $I(1)$. It appears that the results in this paper can be extended to the case when the factors are $I(d)$ $(d = 2, 3, \ldots)$; but, we do not consider such extensions in this paper.

Let $X = \left[X_{it}\right]_{T\times N}$ denote the panel data in matrix form, $F = (F_1, \ldots, F_T)'$ denote the $T \times r$ matrix of unobservable common factors, $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ denote the matrix of factor loadings, and $e = \left[e_{it}\right]_{T\times N}$ denote the matrix of error terms from the factor model. Then the factor model (1) can also be expressed as $X = F\Lambda' + e$. Since the stationary and nonstationary terms need to be treated differently, let us write $F_t' = (E_t', G_t')'$, where E_t is $p \times 1$ and nonstationary, G_t is $q \times 1$ and stationary; p and q are assumed known. Therefore, $E_t = E_{t-1} + u_t$, where u_t is stationary. Substituting $F_t' = (E_t', G_t')'$, the factor model (1) and the FAR model (2) take the forms

$$X_{it} = \lambda_i^{(1)'} E_t + \lambda_i^{(2)'} G_t + e_{it} \qquad (i = 1, ..., N; t = 1, ..., T) \tag{3}$$

$$Y_{t+h} = \alpha' E_t + \beta' G_t + \omega' W_t + \epsilon_{t+h} \qquad (t = 1, ..., T), \tag{4}$$

respectively, where $\lambda_i = (\lambda_i^{(1)'}, \lambda_i^{(2)'})'$ and $\theta = (\alpha', \beta')'$. As expected, estimates of the coefficients α and β of the nonstationary and stationary variables in the FAR model (2), converge at the rates T and $T^{1/2}$ respectively. Similarly, since W_t is $I(1)$, we would expect that the estimator $\hat{\omega}$ to converge at the rate T. To state the results with such different rates of convergence, we introduce the following scaling matrices:

$$D_{1T} = \text{diag}(TI_p,\ T^{1/2} I_q)_{r\times r}, \qquad D_{2T} = TI_m, \qquad D_T = \text{diag}(D_{1T},\ D_{2T}). \tag{5}$$

For a given matrix A, let $A > 0$ denote that it is positive definite. For given matrices X and Y, let $X \oplus Y$ denote $diag(X, Y)$. Finally, let $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and in distribution, respectively.

2.2 Estimation of the common factors

To estimate the latent factors for a given panel dataset X, we may use either the Gaussian Maximum Likelihood Estimator (MLE) or the method based on Principal Component Analysis [PCA]. In this paper, we use the latter. To choose an optimal number of factors, r, we use the Integrated Panel Criterion [IPC] and the panel Information Criterion [IC] introduced by Bai (2004) and Bai & Ng (2002), respectively.

Let $\tilde{F}_{T\times r} = (\tilde{F}_1, \ldots, \tilde{F}_T)'$ be defined as equal to D_{1T} times the matrix formed by the r eigenvectors corresponding to the r largest eigenvalues of the matrix XX'. Since we use PCA, $\tilde{F}$ is an estimator of F, the matrix of common factors. For the derivations of the asymptotic results, we assume that the numbers of stationary and nonstationary factors is known. However, in empirical studies, we apply one or more tests to each factor to determine whether it is stationary or nonstationary. Therefore, it is clear that the estimation method has an element of pre-testing. Assuming that a consistent test is applied for classifying a variable as stationary or nonstationary, it follows that the probability of misclassification tends to zero, and hence the asymptotic results would be unaffected by the pre-testing.

Once the factors have been estimated, a corresponding estimator of the factor loading matrix Λ is $\tilde{\Lambda} = X'\tilde{F}D_{1T}^{-2}$. Without loss of generality, we assume that the columns of $\tilde{F}$ are arranged such that the first p have been classified as nonstationary and their corresponding eigenvalues are in the decreasing order, and the remaining q columns have been classified as stationary and their corresponding eigenvalues are in the decreasing order. Therefore, without loss of generality, we write $\tilde{F}_t = (\tilde{E}_t', \tilde{G}_t')'$ and $F_t = (E_t', G_t')'$. Let $\tilde{V}_{p,NT}$ denote the diagonal matrix with diagonal elements equal to the largest p eigenvalues of XX' divided by T^2N and each of the corresponding eigenvector has been classified as nonstationary; further, without loss of generality, assume that the diagonal elements appear in the decreasing order. Similarly, let $\tilde{V}_{q,NT}$ denote the diagonal matrix with diagonal elements equal to the largest q eigenvalues of XX' divided by TN and each of the corresponding eigenvector has been classified as stationary; again, without loss of generality, assume that the diagonal elements appear in the decreasing order. Let $\tilde{V}_{NT} = diag(\tilde{V}_{p,NT}, \tilde{V}_{q,NT})$. Therefore, $\tilde{V}_{NT}$ is equal to the diagonal matrix whose diagonal elements are the $r = (p+q)$ largest eigenvalues of the matrix XX' multiplied by D_{1T}^{-2}/N.

We adopt the standard procedure to ensure that the factors are identified up to a rotation. To this end, we assume that $\tilde{F}$ satisfies the normalization $D_{1T}^{-2}\tilde{F}'\tilde{F} = I_r$, $\tilde{\Lambda}'\tilde{\Lambda}$ is diagonal, and define the rotation matrix $H = N^{-1}\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda$. If all the variables are stationary then the foregoing H reduces to the expression in Bai & Ng (2002), and if all the variables are nonstationary then it reduces to the forms in Bai (2004) and Choi (2017).

Let $\hat{L}_t = (\tilde{F}_t', W_t)'$ and $\delta = (\theta' H^{-1}, \omega')'$; then, H and δ are also a functions of the data and unknown population parameters. Then, the FAR model (4) can be written as

$$\begin{aligned} Y_{t+h} &= \theta' F_t + \omega' W_t + \epsilon_{t+h} = \theta' H^{-1}(HF_t - \tilde{F}_t + \tilde{F}_t) + \omega' W_t + \epsilon_{t+h} \\ &= \theta' H^{-1}\tilde{F}_t + \omega' W_t + \theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h} \\ &= \delta' \hat{L}_t + \theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h}. \end{aligned} \tag{6}$$

Let $(\hat{\alpha}', \hat{\beta}', \hat{\omega}')$ denote the the ordinary least squares [OLS] estimator of $(\alpha', \beta', \omega')$ obtained by regressing Y_{t+h} on $\hat{L}_t$ $(t = 1, \ldots, T-h)$. Then

$$\hat{\delta} = (\hat{\alpha}', \hat{\beta}', \hat{\omega}')' = \left(\sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t{}'\right)^{-1} \sum_{t=1}^{T-h} \hat{L}_t Y_{t+h}. \tag{7}$$

Later we will show that $\theta' H^{-1}(HF_t - \tilde{F}_t)$ in (6) is asymptotically centered at zero in the limit, and hence $\{\theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h}\}$ could be treated as an error term centered at zero for the purposes of estimating δ. In consequence, it turns out that $\hat{\delta} - \delta$ is asymptotically normal with mean zero, which will be used later for deriving a prediction interval.

Remark: While it is not essential for the derivations, the following observation is helpful for interpretation. The stationary and nonstationary terms behave as if they are independent, and the rotation H can be performed separately for the stationary and nonstationary terms. To this end, we may define $H_1 = \tilde{V}_{p,NT}^{-1} \frac{\tilde{E}'E}{T^2} \frac{\Lambda_1' \Lambda_1}{N}$, $H_2 = \tilde{V}_{q,NT}^{-1} \frac{\tilde{G}'G}{T} \frac{\Lambda_2' \Lambda_2}{N}$, and $H_0 = diag(H_1, H_2)$. Then $(H - H_0)$ converges in probability to zero. Consequently, for the asymptotic results, the rotation of the entire factor by H leads to the same results as performing the rotations separately for the nonstationary and stationary factors by H_1 and H_2, respectively.

2.3 Distribution theory

In this section, we study the asymptotic distributions of the generated factors and the estimators of the regression parameters. First, we introduce some assumptions; in these assumptions, $M \in \mathbb{R}$ denotes a generic constant, and hence it may be different in its different appearances.

Assumption 1 [Factors and factor loadings]. (i) The strictly stationary process u_t in $E_t = E_{t-1} + u_t$, satisfies $\max_{t \geq 1} E \|u_t\|^{4+\delta} \leqslant M$, for some $\delta > 0$. (ii) $E \|F_1\|^4 \leqslant M$ and $D_{1T}^{-1} \sum_{t=1}^{T} F_t F_t' D_{1T}^{-1} \xrightarrow{d} \Sigma_F$ as $T \to \infty$, where Σ_F is a positive definite random matrix. (iii) The number of factors r is known and does not depend on N or T; further, factors are not cointegrated. (iv) The loadings λ_i are either deterministic and $\|\lambda_i\| \leqslant M$ satisfying $\Lambda'\Lambda/N \to \Sigma_\Lambda$ as $N \to \infty$, or they are stochastic and $E \|\lambda_i\|^4 \leqslant M$ satisfying $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda$ as $N \to \infty$, for some $r \times r$ positive definite non-random matrix Σ_Λ. (v) The eigenvalues of the matrix $\Sigma_\Lambda \Sigma_F$ are distinct, almost surely.

To estimate the number of factors, we assume that the factors are not cointegrated. If they are cointegrated then the stationary and nonstationary factors cannot be identified because one $I(0)$ factor may represent a combination of cointegrated $I(1)$ factors. By assuming Σ_F and Σ_Λ are positive definite and the eigenvalues of $\Sigma_\Lambda \Sigma_F$ are distinct, we ensure the identifiability of the r factors. If all the factors are nonstationary then Σ_F is distributed as $\int_0^1 B_F(r)B_F'(r)dr$, and if the factors are all stationary then Σ_F converges to the variance-covariance matrix of the factors. To state the next assumption, let us introduce the following notation:

$$\gamma_{st} = E\left(N^{-1}\sum_{i=1}^{N} e_{is}e_{it}\right),\ \tau_{ij,t} = E\left(e_{it}e_{jt}\right),\ \tau_{ij,ts} = E\left(e_{it}e_{js}\right)\ \ (i,j=1,\ldots,N;\ s,t=1,\ldots,T).$$

Assumption 2 [Idiosyncratic errors]. (i) $E\left(e_{it}\right)=0$ and $E\left|e_{it}\right|^8 \leqslant M$ $(i=1,\ldots,N;t=1,\ldots,T)$. (ii) $|\gamma_{ss}| \leqslant M$ $(s=1,...,T)$, and $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{st}| \leqslant M$. (iii) $|\tau_{ij,t}| \leqslant |\tau_{ij}|$, for some τ_{ij} $(i,j=1,\ldots,N;\ t=1,\ldots,T)$, and $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}| \leqslant M$. (iv) $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \leqslant M$. (v) $E\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{is}e_{it}-E\left(e_{is}e_{it}\right)\right]\right|^4 \leqslant M$ $(t,s=1,\ldots,T)$.

Assumption 2 allows the idiosyncratic errors to have weak serial and cross sectional dependence. Heteroskedasticity is also allowed in both the serial and the cross-section dimensions. Since we allow weak correlations among the idiosyncratic errors in (1), it is an *approximate factor model*; for simplicity, we refer to it simply as a *factor model*.

Assumption 3. [Dependence among $\lambda_i, F_t,$ and e_{it}]. (i) $E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|D_{1T}^{-1}\sum_{t=1}^{T}F_te_{it}\right\|^2\right) \leqslant M$, and $E(F_t\, e_{it})=0$ $(i=1,\ldots,N;\ t=1,\ldots,T)$. (ii) $(1/N)\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\lambda_i\lambda_j' e_{it}e_{jt}\right) \to \Gamma_t$ as $N\to\infty$, for some Γ_t, and $N^{-1/2}\Lambda' e_t \xrightarrow{d} N\left(0,\Gamma_t\right)$ as $N\to\infty$, for each fixed t $(t=1,\ldots,T)$. (iii) $E\left\|N^{-1/2}D_{1T}^{-1}\sum_{t=1}^{T}\Lambda' e_tF_t'\right\|^2 \leqslant M$.

Assumption 3 allows the factor loadings $\{\lambda_1,\ldots,\lambda_N\}$, the factors $\{F_1,\ldots,F_T\}$, and the idiosyncratic errors $\{e_{it}, i=1,\ldots,N;t=1,\ldots,T\}$ to have a weak dependence among them.

2.3.1 Consistency of the generated factors

In the literature on FAR models, the consistency of the generated factors has been established for both stationary and nonstationary factors separately. Bai & Ng (2002) and Bai (2004) showed that the time-averaged mean square of factor estimation error [MSE] has $\min\{N,T\}$ and $\min\{N,T^2\}$ convergence rates for $I(0)$ and $I(1)$ factors separately. In our setting, the set of latent factors F contains a mixture of $I(1)$ and $I(0)$ series, and we show that the generated factors are jointly consistent and the convergence rate of MSE is

$\min\{N, ||D_{1T}^{-2}||^{-1}\}$. To state the consistency of generated factors in the next lemma, let us recall that the rotation matrix H was defined as $N^{-1}\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda$.

Lemma 1. *Suppose that Assumptions 1-3 are satisfied. Let* $\delta_{NT}^{-1} = \max\left[N^{-1/2}, ||D_{1T}^{-1}||\right]$*. Then,* $T^{-1}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$.

This lemma states that the time averaged square of factor estimation error converges to zero as $N, T \to \infty$ and the convergence rate is $\min\{N, ||D_{1T}^{-2}||^{-1}\}$. Therefore, we may estimate a rotation of the mixture of latent factors consistently by the method of principal component analysis. For the case when all the factors are stationary, the scaling matrix D_{1T} is $\sqrt{T}I_r$ and the convergence rate is $\min\{N, T\}$; this is consistent with the corresponding result in Bai (2003). For the case when all the factors are nonstationary, $D_{1T} = TI_r$ and the convergence rate is $\min\{N, T^2\}$; this is consistent with Bai (2004).

2.3.2 Asymptotic distribution of the generated factors

To derive the asymptotic distributions of the estimated factors, we introduce the following additional assumption.

Assumption 4 [Weak dependence of idiosyncratic errors]. (i) $\sum_{s=1}^{T}|\gamma_{st}| \leqslant M$ $(t = 1, ..., T)$, and (ii) $\sum_{j=1}^{N}|\tau_{ij}| \leqslant M$ $(i = 1, ..., N)$, where γ_{st} and τ_{ij} are as in Assumption 2.

Lemma 2. *Suppose that Assumptions 1-4 are satisfied. Then* $D_{1T}^{-2}\tilde{F}'F \xrightarrow{d} Q$ *as* $N, T \to \infty$*, where* $Q = V^{1/2}\Upsilon'\Sigma_{\Lambda}^{-1/2}$ *is a random matrix,* $V = diag(v_1, ..., v_r)$ *with* $\{v_1, \ldots, v_r\}$ *denoting the eigenvalues of* $\Sigma_\Lambda\Sigma_F$*, and* Υ *is the corresponding matrix formed by scaled eigenvectors such that* $\Upsilon'\Upsilon = I_r$.

Lemma 3. *Suppose that Assumptions 1– 4 hold. Then, as* $N, T \to \infty$ *with* $\sqrt{N}||D_{1T}^{-2}|| \to 0$*, for each* t*, we have* $\sqrt{N}\left(\tilde{F}_t - HF_t\right) \xrightarrow{d} V^{-1}QN\left(0, \Gamma_t\right) \stackrel{d}{=} N(0, \Sigma_{\tilde{F}})$*, where* Q *is defined in Lemma 2,* Γ_t *is defined in Assumption 3, and* Q *is independent of* $N(0, \Gamma_t)$.

This lemma shows that the factor estimation error is asymptotically normal with mean zero; this is important for estimating the parameters of the FAR model consistently, as indicated previously. Later we show that the asymptotic variance of $\left(\tilde{F}_t - HF_t\right)$ can be estimated consistently by $\tilde{V}_{NT}^{-1}\hat{\Gamma}_t\tilde{V}_{NT}^{-1}$, which is used for constructing the prediction interval of h-step ahead forecasts.

2.3.3 Asymptotic distribution of the estimators

To obtain the asymptotic distribution of the OLS estimator $\hat{\delta}$ of δ, we introduce the following additional assumptions.

Assumption 5 [Weak dependence between idiosyncratic and regression errors.].

i. $E\left|(TN)^{-1/2}\sum_{s=1}^{T-h}\sum_{i=1}^{N}\{e_{is}e_{it}-E\left(e_{is}e_{it}\right)\}\epsilon_{s+h}\right|^2 \leqslant M \quad (t=1,\ldots,T; h>0)$.

ii. $E\left\|(TN)^{-1/2}\sum_{t=1}^{T-h}\sum_{i=1}^{N}\lambda_i e_{it}\epsilon_{t+h}\right\|^2 \leqslant M$, and $E\left(\lambda_i e_{it}\epsilon_{t+h}\right)=0 \quad (i=1,\ldots,N; t=1,\ldots,T)$.

Assumption 6 [Moment and CLT for score vector.]. Let $L_t=(F_t', W_t')'$. Then, the following conditions are satisfied. (i) $E\left(\epsilon_{t+h}\right)=0$ and $E\left|\epsilon_{t+h}\right|^2 < M$ $(t=1,\ldots,T)$.
(ii) $D_T^{-1}\sum_{t=1}^{T}L_tL_t'D_T^{-1} \xrightarrow{d} \Sigma_L$ as $N,T\rightarrow\infty$, where Σ_L is a positive definite random matrix. (iii) $D_T^{-1}\sum_{t=1}^{T}L_t\varepsilon_{t+h} \xrightarrow{d} \Sigma_{\epsilon L}^{1/2}\ N(0,I)$, where $\Sigma_{\epsilon L}>0$ with probability one.

Assumption 5 imposes restrictions on the degree of dependence among the idiosyncratic errors over time, and between the idiosyncratic and regression errors. Part (ii) of Assumption 5 holds if $\{\lambda_i\},\{e_{it}\}$, and $\{\epsilon_t\}$ are mutually independent and Assumption 2 holds.

Theorem 1. *Suppose that Assumptions 1-6 hold and that* $T/N\rightarrow 0$. *Let* δ *and the OLS estimator* $\hat{\delta}$ *be as in (6) and (7), respectively. Then, as* $(N,T)\rightarrow\infty$, *we have* $D_T(\hat{\delta}-\delta)\xrightarrow{d} N\big(0,\Sigma_\delta\big)$ *and* $H\oplus I\xrightarrow{d}\Psi$, *where* $\Sigma_\delta=(\Psi')^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}\Sigma_L^{-1}\Psi^{-1}$, *and* Σ_L *and* $\Sigma_{\epsilon L}$ *are as defined in Assumptions 1-6.*

The appearance of the scaling matrix $D_T=diag(TI_p, T^{1/2}I_q, TI_m)$ in Theorem 1 shows that the estimators $\hat{\alpha},\hat{\beta}$, and $\hat{\omega}$ converge at the rates $T, T^{1/2}$, and T respectively. Consequently, the limiting distribution in this theorem reduces to the following known corresponding results in: (a) Bai & Ng (2002) for the FAR model with $I(0)$ variables only, and (b) Choi (2017) for the FAR model with $I(1)$ variables only.

Since the limiting normal distribution in Theorem 1 has mean zero, it follows that the use of generated factors, instead of the original unobservable factors in the model, does not affect the consistency of the estimators. To arrive at this result, we used the assumption $T/N\rightarrow 0$, which ensures that the effect of the error resulting from factor estimation becomes negligible in the limit. By contrast, if the assumption $T/N \to 0$ is replaced by $T/N \to c$, for some $0 < c < \infty$, then the limiting normal distribution would have a nonzero mean, and hence the estimator would not be consistent. In fact, Gonçalves & Perron (2014) showed, for the case when all the variables are $I(0)$, that if $\sqrt{T}/N \to c$ for some $0 < c < \infty$, then there would be asymptotic bias.

The unknown covariance matrix Σ_δ may be estimated consistently by

$$\hat{\Sigma}_{\hat{\delta}} = \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1} \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{\epsilon}_{t+h}^2 \hat{L}_t \hat{L}_t' D_T^{-1} \right) \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1} . \quad (8)$$

This estimator is robust against heteroskedasticity in the regression error. For the special case of homoskedastic errors, a simpler consistent estimator of Σ_δ is

$$\hat{\Sigma}_{\hat{\delta}} = \hat{\sigma}_\epsilon^2 \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1} , \quad (9)$$

where $\hat{\sigma}_\epsilon^2 = T^{-1} \sum_{t=1}^{T-h} \hat{\epsilon}_{t+h}^2$ is an estimator of the variance of regression errors.

2.4 Prediction interval

Let $Y_{T+h|T}$ denote the conditional mean $E[Y_{T+h} \mid \mathcal{F}_T]$ where $\mathcal{F}_T$ is the information up to time T, and let $(\hat{\delta}, \hat{L}_t)$ be as in (6) and (7). Then, an estimator of $Y_{T+h|T}$ is $\hat{Y}_{T+h|T} = \hat{\delta}' \hat{L}_T$; similarly, $\hat{Y}_{T+h} = \hat{\delta}' \hat{L}_T$ is also a point forecast of Y_{T+h}. In this section, we obtain a confidence interval for $Y_{T+h|T}$ and a prediction interval for Y_{T+h}. These are obtained by using the next theorem.

Theorem 2. *Suppose that Assumptions 1–6 hold. Further, suppose also that* $\sqrt{N}||D_{1T}^{-2}|| \to 0$ *and* $T/N \to 0$ *as* $N, T \to \infty$, *and that* $(\hat{\Sigma}_\delta, \hat{\Sigma}_{\tilde{F}})$ *is a given consistent estimator of* $(\Sigma_\delta, \Sigma_{\tilde{F}})$. *Then, we have* $\hat{B}_T^{-1/2}\{\hat{Y}_{T+h|T} - Y_{T+h|T}\} \xrightarrow{d} N(0,1)$ *as* $N, T \to \infty$, *where* $\hat{B}_T = \left[\hat{L}_T D_T^{-1} \hat{\Sigma}_\delta D_T^{-1} \hat{L}_T' + N^{-1} \hat{\theta}' \hat{\Sigma}_{\tilde{F}} \hat{\theta} \right]$ *is a consistent estimator of the asymptotic variance, denoted* B_T, *of the conditional forecasting error that appears in the numerator.*

To provide some insight into the foregoing suggested form for $\hat{B}_T$, note that the forecast error can be expressed as

$$\hat{Y}_{T+h|T} - Y_{T+h|T} = (\hat{\delta} - \delta)' \hat{L}_T + \theta' H^{-1} (\tilde{F}_T - H F_T). \quad (10)$$

This forecast error is the sum of two components: the first is due to the error in estimating δ and the other is due to estimating the factor. Theorem 1 and Lemma 3 show that each of these is asymptotically normal with mean zero. It turns out that these two are essentially asymptotically independent and hence the asymptotic variances simply add up.

To use Theorem 2 for inference in empirical studies, we need a suitable consistent estimator $(\hat{\Sigma}_\delta, \hat{\Sigma}_{\tilde{F}})$ of $(\Sigma_\delta, \Sigma_{\tilde{F}})$. For, $\hat{\Sigma}_\delta$, we may use the estimators in (8) or (9) depending on the assumptions. Using Lemmas 2 and 3, a consistent estimator of $\Sigma_{\tilde{F}_T}$ is

$$\hat{\Sigma}_{\tilde{F}_T} = \tilde{V}_{NT}^{-1} \hat{\Gamma}_T \tilde{V}_{NT}^{-1}, \tag{11}$$

where $\hat{\Gamma}_T$ is an estimator of the asymptotic covariance matrix of $\left(N^{-1/2}\Lambda' e_t\right)$, and $\tilde{V}_{NT}$ was defined as a diagonal matrix of the largest r eigenvalues of XX' multiplied by $D_{1T}^{-2}N^{-1}$.

To make use of the form in (11), we need a feasible estimator of Γ_T. As suggested by Bai & Ng (2006), depending on the assumptions, $\hat{\Gamma}_T$ may take one of the following forms:

$$(a)\ \hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N} \hat{e}_{iT}^2 \tilde{\lambda}_i \tilde{\lambda}_i', \quad (b)\ \hat{\Gamma}_T = \hat{\sigma}_e^2 \frac{1}{N}\sum_{i=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_i', \quad (c)\ \hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_j' \hat{e}_{iT}\hat{e}_{jt}, \tag{12}$$

where $\hat{e}_{it} = X_{it} - \tilde{\lambda}_i' \tilde{F}_t$. For cross sectionally uncorrelated idiosyncratic errors, the first two forms of $\hat{\Gamma}_T$ in (12) are suitable. If the errors are homoskedastic and $E(e_{it}^2) = \sigma_e^2$, say, then σ_e^2 can be estimated by $\hat{\sigma}_e^2 = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2$ and the second form in (12) would be suitable. The third form in (12) is suitable for estimating the asymptotic variance of generated factors when the idiosyncratic errors have cross sectional correlation. By combining the aforementioned estimators, we obtain a feasible estimator $\hat{B}_T$. Using these, a $100(1-\alpha)\%$ confidence interval for the conditional mean $Y_{T+h|T}$ is

$$\left(\hat{Y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B}_T} \quad , \quad \hat{Y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B}_T}\right), \tag{13}$$

where $z_{1-\alpha/2}$ stands for $(1-\alpha/2)th$ quantile of the standard normal distribution.

Next, consider constructing a forecast interval for Y_{T+h}. To this end, first note that the forecast error is

$$\hat{\epsilon}_{T+h} = \hat{Y}_{T+h|T} - Y_{T+h} = \hat{L}_T'\left(\hat{\delta} - \delta\right) + \theta' H^{-1}\left(\tilde{F}_T - HF_T\right) - \epsilon_{T+h}. \tag{14}$$

Therefore, the limiting distribution of forecast error also depends on the distribution of the regression error ϵ_{T+h}. Let us suppose that $\epsilon_{T+h} \sim N(0, \sigma_\epsilon^2)$. Then, it follows from Theorem

2 that the forecasting error $\hat{\epsilon}_{T+h}$ is also asymptotically normal with mean zero and variance $B_T + var(\epsilon)$. Let $\hat{\sigma}_\epsilon^2$ denote a consistent estimator of σ_ϵ^2; for example, if $\{\varepsilon_t\}$ are iid, then we may choose $\hat{\sigma}_\epsilon^2 = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$. Then, an asymptotic 95% prediction interval for Y_{T+h} is

$$\left(\hat{Y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B}_T + \hat{\sigma}_\epsilon^2} \quad , \quad \hat{Y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B}_T + \hat{\sigma}_\epsilon^2}\right). \tag{15}$$

3 Simulation study: finite sample properties

Design of the simulation study

The data generating process [DGP] for the FAR is

$$Y_{t+1} = \alpha F_{1t} + \beta F_{2t} + \omega Y_t + \epsilon_{t+1} \qquad (t = 1, \ldots, T-1) \tag{16}$$

$$F_{1t} = F_{1,t-1} + v_t, \qquad (v_t, F_{2,t}) \sim MVN(0, C), \quad C = (1, \rho \mid \rho, 1). \tag{17}$$

For ρ in (17), we considered the values $0.0, 0.5$, and 0.9. For the error term ϵ_t, we considered both homoskedastic and heteroskedastic cases - see below. The $T \times N$ panel data set was generated by

$$X_{it} = \lambda_i^{(1)} F_{1t} + \lambda_i^{(2)} F_{2t} + e_{it}, \tag{18}$$

with the λ_i's drawn from $N(0,1)$ and the error terms $\{e_{it}\}$ as stated below.

Sixteen combinations of $[T, N]$ were considered with $T = 30, 50, 100, 200$ and $N = 30, 50, 100, 200$. The parameter values were set at $\alpha = 0.5, \beta = 1$, and $\omega = 0.5$. We considered the following three different DGPs for each ρ: (1) DGP1: $e_{it} \sim N(0,1)$ and $\epsilon_t \sim N(0,1)$; (2) DGP2: $e_{it} \sim N(0,1)$ and $\epsilon_t \sim N\left(0, 3^{-1}F_{2t}^2\right)$; (3) DGP3: $e_{it} \sim N(0, \sigma_i^2)$ and $\epsilon_t \sim N\left(0, 3^{-1}F_{2t}^2\right)$.

Among the three DGP's, DGP1 is the simplest for which the errors are i.i.d. in both time and cross-section dimensions. In DGP2, $var(\epsilon_t)$ depends on the stationary factor, and hence conditionally heteroskedastic over time. In DGP3, $var(\epsilon_t)$ varies over time and $var(e_{it})$ is distributed uniformly over $[0.5, 1.5]$; therefore, the average variance is the same as that for the homoskedastic case. All simulation estimates are based on 5000 repeated samples. Since the FAR model has a lag term, we adopted a burn-in period of 100 time units; thus, for generating each sample, the first 100 observations were discarded. We used the $\hat{\Sigma}_{\hat{\delta}}$ in (9) and (8) for DGP1 and {DGP2, DGP3}, respectively.

The results are reported in two parts. The first part reports the simulation results for the coverage rates of 95% prediction intervals in (15). The second part compares the out-of-sample forecast performance of the method based on the mixture-FAR model developed in this paper with the methods based on a nonstationary-FAR and the AR(4) models.

3.1 Coverage rates of prediction intervals

Table 1 reports the coverage rates of 95% prediction intervals for Y_{T+1}. These are based on the assumption that the regression errors are normal. The coverage rates of these intervals range from 88% to 98% with most of them being close to the nominal 95%. Therefore, in terms of coverage rates, the prediction intervals performed quite well.

Table 1: Coverage rates (%) of 95% prediction intervals for one-step ahead forecasts.

		$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
	T\N	30	50	100	200	30	50	100	200	30	50	100	200
DGP1	30	93	93	91	92	94	94	92	92	94	93	92	91
	50	91	94	93	90	92	94	92	89	92	94	92	90
	100	90	89	91	88	92	90	91	89	91	90	91	88
	200	92	90	89	90	92	90	89	89	93	89	90	89
DGP2	30	97	96	95	94	95	95	95	94	97	98	97	95
	50	96	96	98	94	97	97	94	94	98	97	96	95
	100	97	95	97	95	96	95	95	95	96	96	95	95
	200	96	96	96	94	96	96	94	94	97	96	95	94
DGP3	30	97	97	96	94	97	96	96	95	97	98	95	96
	50	97	95	96	96	97	95	96	96	98	96	96	95
	100	97	96	95	97	96	96	96	95	96	95	96	96
	200	95	97	95	96	95	95	95	95	96	94	95	96

Note: The assumed error distribution of the forecasting model is normal.

3.2 Performance of mixture-FAR relative to non-stationary FAR

In this subsection, we evaluate the performance of the mixture-FAR method relative to nonstationary-FAR method. Recall that the nonstationary-FAR model (see Choi 2017), requires all the variables in the FAR model to be $I(1)$. We evaluate the out-of-sample forecast performance in terms of *out-of-sample R-square*, denoted R^2_{os}, defined as

$$R^2_{os} = 1 - \left(\sum_{i=T_1+1}^{T} (Y_i - \hat{Y}_i)^2 \right) \left(\sum_{i=T_1+1}^{T} (Y_i - \tilde{Y}_i)^2 \right)^{-1}, \tag{19}$$

where $\hat{Y}_i$ = prediction using the mixture-FAR model, $\tilde{Y}_i$ = prediction using the competing or reference model, the observations from the first $(T_1 + j)$ time points are used for estimating the model, and the observation at time $T_1 + j + 1$ is used for evaluating the performance of the out-of-sample forecast at time $(T_1 + j + 1)$ $(j = 0, \ldots, T - (T_1 + 1))$. Thus, R^2_{os} is a measure of how well the mixture-FAR performed during the period $[T_1+1, T]$, relative to the competing model. As an example, if $R^2_{os} = 0.1$ (respectively, $R^2_{os} = -0.1$) then an estimate of the MSE of prediction for the mixture-FAR model is 10% lower (respectively, higher) than that for the competing model. In this simulation study, we chose the nonstationary-FAR as the competing model. Throughout this paper, forecast evaluations are based on expanding windows for the estimation period, unless the contrary is made clear.

In this part of the study, we considered the DGP1 with $T = [60, 90, 150, 300]$ and $T_1 = [40, 60, 100, 200]$. First, we consider forecasting a nonstationary series Y_t using the mixture-FAR model and compare it with the corresponding nonstationary-FAR. Table 2 provides the results for this comparison. It is evident that the mixture-FAR method performed significantly better than the competing nonstationary-FAR in terms of R^2_{os}. As an example, the entry 0.43 in the cell for $T_1 = 40$ and $N = 30$ shows that the MSE of prediction for the mixture-FAR model is 43% lower than that for the nonstationary-FAR model. The table also shows that, for every case considered in Table 2, the MSE of prediction for the proposed mixture-FAR model is at least 33% lower than that for the nonstationary-FAR model. Therefore, in this simulation study, the improvement of the mixture-FAR model compared to the nonstaionary-FAR model is substantial.

In summary, for every case that we studied, the mixture-FAR model performed significantly better than its corresponding competitor, the nonstationary-FAR, for forecasting a nonstationary variable.

Table 2: Values of R^2_{os} for the performance of mixture-FAR model relative to the corresponding nonstationary-FAR model.

$T_1 \backslash N$	$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
	30	50	100	200	30	50	100	200	30	50	100	200
40	0.43	0.44	0.45	0.46	0.41	0.42	0.42	0.43	0.33	0.34	0.35	0.35
60	0.45	0.46	0.46	0.47	0.42	0.43	0.44	0.45	0.36	0.36	0.36	0.37
100	0.46	0.47	0.48	0.48	0.44	0.45	0.46	0.46	0.37	0.38	0.38	0.39
200	0.47	0.48	0.48	0.49	0.45	0.46	0.46	0.46	0.38	0.39	0.39	0.39

Note: The values in this table are for one-step ahead out-of-sample forecasts. The forecasting variable Y is nonstationary [$I(1)$].

4 Empirical application

In this section we apply the mixture-FAR model for forecasting two key non-stationary macroeconomic variables, namely the GDP and the Industrial Production [IP]. Since we use quarterly data, we start with a basic AR(4) model and augment it with factors to construct FAR models. For each model, we compute two sets of prediction intervals, one is based on the asymptotic distribution of the standardized forecast and the other is based on the t-percentile bootstrap; the validity of the bootstrap is yet to be established.

Recall that the forecast of the conditional mean, as shown in Theorem 2, is asymptotically normal; this result does not require that the functional form of the error distribution be known. The indications are that a residual based bootstrap method is likely to be valid for constructing confidence interval for the forecast conditional mean and for constructing a prediction interval. With this in mind, we expanded the simulation study in the previous section and evaluated the coverage rates of residual based t-percentile bootstrap prediction intervals when the error distribution is normal and when it is t with 5 degrees of freedom. Since we used residual based bootstrap, it does not assume that the error distribution is known. The results are presented in the two tables at the end of the Supplement; they show that the coverage rates of the bootstrap prediction intervals are close to the nominal level. Therefore, the indications are that it is reasonable to compare the bootstrap intervals with those based on (15). We compare and contrast the out-of-sample forecasting performance of the mixture-FAR model with the corresponding non-stationary-FAR and the AR(4) models. To quantify the out-of-sample forecasting performance, we use R^2_{os} defined in (19).

4.1 Data description

The data were collected from FRED-MD and FRED-QD; these are well-known databases for macroeconomic variables containing monthly and quarterly data, respectively. The latter contains 246 US macroeconomic time series for the period 1959:Q1 to 2018:Q4, with a total of 240 (T=240) observations. We excluded 36 variables because there were missing observations, and used a balanced panel for 210 variables. The variables are categorized into 14 groups; for more details, see the updated appendix for FRED-QD at $https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QDappendix.pdf$. The macroeconomic variables in this balanced panel data set are further categorized into two levels of aggregation, 110 "high-aggregates" and 100 "sub-aggregates". The panel data for $N = 100$ sub-aggregates were used for estimating the factors; to this end, we used principal components analysis [PCA]. These sub-aggregates consist of both stationary and nonstationary time series; for each series, the transformation to $I(0)$ is given in the third row of the data set.

4.2 Estimation of factors

We adapted the methods proposed in Bai & Ng (2002) and Bai (2004) for choosing an 'optimal' number of factors. The method proposed by Bai & Ng (2002), which is based on information criteria, led to the total number of factors being eight for the set of 100 sub-aggregate macroeconomic variables. Then, we applied the method based on *iterated panel criterion* proposed by Bai (2004), and concluded that four of the eight factors are $I(1)$, and the other four are $I(0)$. Finally, we applied the Augmented-Dickey Fuller [ADF] test to each of the factors, and observed that if the factors are ordered according to the magnitude of the eigenvalues, then the factors $\{1, 2, 4, 5\}$ are $I(1)$ and the remaining ones, namely $\{3, 6, 7, 8\}$, are $I(0)$. The overall trends exhibited by the factors in Figure 1 are consistent with the aforementioned observation that the factors $\{1, 2, 4, 5\}$ are $I(1)$ and the other four are $I(0)$. Figure 1 shows time series plots of the estimated factors. Plots of the two high-aggregate macroeconomic variables, GDP and IP, are presented in Figure 2. This figure shows that both variables are nonstationary. Therefore, the mixture-FAR method developed in this paper is potentially applicable for forecasting GDP and IP.

For the data set in this empirical study, the panel data model and the forecasting model with a mixture of stationary and nonstationary factors take the forms,

$$X_{it} = \lambda_i' F_t + e_{it} = \lambda_i^{(1)\prime} E_t + \lambda_i^{(2)\prime} G_t + e_{it},$$
$$Y_{t+h} = \alpha' \tilde{E}_t + \beta' \tilde{G}_t + \omega_1 Y_t + \omega_2 Y_{t-1} + \omega_3 Y_{t-2} + \omega_4 Y_{t-3} + \epsilon_{t+h} \quad (h > 0),$$

where $\tilde{E}_t$ is the set of four nonstationary generated factors, $\tilde{G}_t$ is the set of four stationary generated factors, and $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$ and $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ are their coefficients.

4.2.1 Assessing the out-of-sample forecast performance of mixture-FAR method

We considered the following four models; the basic AR(4), and three mixture-FAR models obtained by augmenting the AR(4) with a mixture of the $I(0)$ and the $I(1)$ factors:

$$\begin{aligned}
\text{Model 1}: Y_{t+h} &= \alpha' \tilde{E}_t + \beta' \tilde{G}_t + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+h}, \\
\text{Model 2}: Y_{t+h} &= \alpha_1' \tilde{E}_t + \alpha_2' \tilde{E}_{t-1} + \beta_1' \tilde{G}_t + \beta_2' \tilde{G}_{t-1} + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+h}, \\
\text{Model 3}: Y_{t+h} &= \sum_{i=0}^{3} \alpha_{1+i}' \tilde{E}_{t-i} + \beta_1' \tilde{G}_t + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+h}, \\
\text{Model 4}: Y_{t+h} &= \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+h}.
\end{aligned}$$

Model 4, the basic AR(4) model, is used as the benchmark for forecast comparison; since we use quarterly data this is a suitable benchmark. Model 1 is the AR(4) model augmented

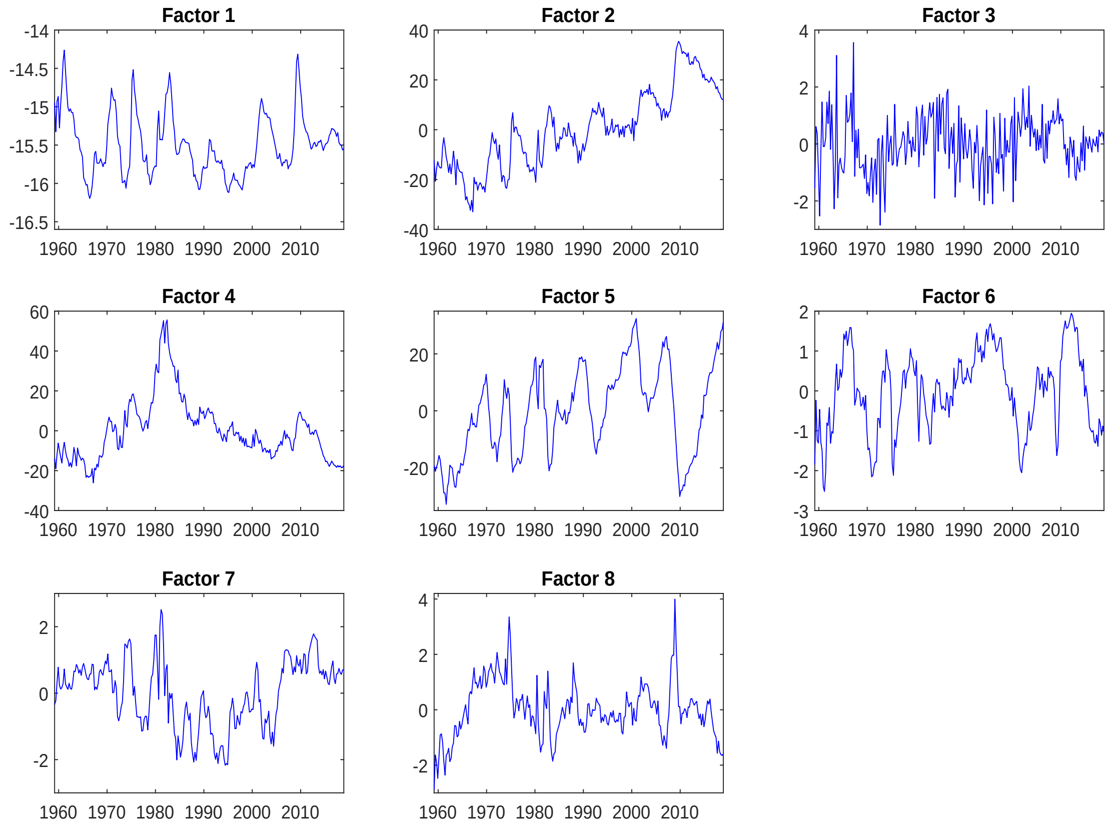

Figure 1: A plot of the eight factors generated from the panel data set of 100 variables

with the eight generated factors; this is a mixture-FAR model. Model 2 is Model 1 augmented with one lag of each generated factor. Model 3 is Model 1 augmented with three lags of each nonstationary factor and the stationary factor with no lags. Thus, Model 1 is nested in Models 2 and 3. Much of this section focuses on comparing and contrasting the out-of-sample forecast performance of Models 1 to 4. The plots in Figures 3 and 4 show the one-step ahead out-of-sample forecasts of $\log(GDP)$ and $\log(IP)$, respectively, for the period 2006:Q1 - 2018:Q4 with the initial estimation period being 1959:Q1 - 2005:Q4. These plots indicate that the out-of-sample predictions of the two I(1) variables, GDP and IP, generally appear to be good.

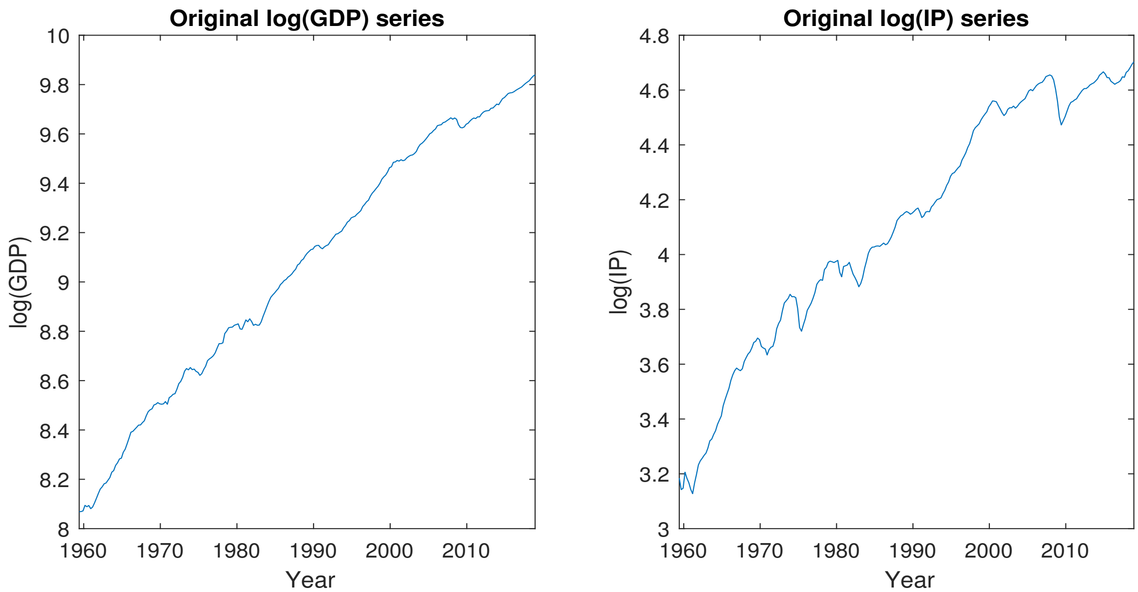

Figure 2: Time series plots of $\log(GDP)$ and $\log(IP)$ for 1959:Q1 – 2018:Q4.

4.2.2 One-step ahead out-of-sample forecast evaluations

We assess the relative predictive performance of the four models in terms of the out-of-sample R^2, denoted R^2_{os}, defined in (19). In this subsection, AR(4) is used as the basic benchmark; later we consider a nonstationary-FAR as the benchmark. We considered three different first estimation periods; we also evaluated the forecasts with a rolling window of 40 years for the estimation period, but there were no improvements in the forecast performance, compared to the expanding window. The values of R^2_{os} in Table 3 indicate that the mixture-FAR model, Model 2, outperforms the benchmark model, AR(4) for forecasting GDP and IP. Overall, the results presented in Table 3 indicate that Model 2 performs better than the other two models as well.

4.3 Forecast evaluations with long forecast horizons

So far, we considered one-step ahead forecasts. Next, we evaluate and compare Models 1, 2, and 3 in terms of the accuracy of their forecasts over longer forecast horizons, instead of just one-step ahead. We computed the forecasts with the first estimation period being 1959:Q1 - 1999:Q4. We calculated R^2_{os} for different specifications of mixture-FAR relative to AR(4). Figures 5 and 6 provide plots of R^2_{os} against the forecast horizon h. These figures show that

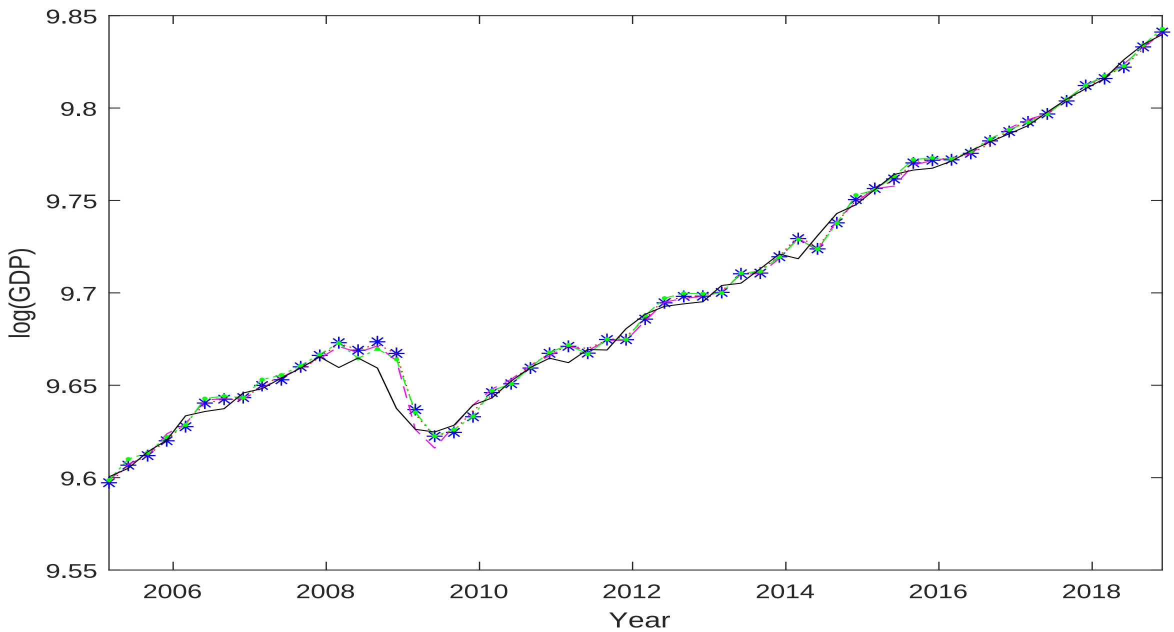

Figure 3: The observed $\log(GDP)$, and plots of one-step ahead out-of-sample forecasts of $\log(GDP)$ for 2006:Q1 – 2018:Q4. Blue * : predicted series with Model 1. Magenta dashed line - -: predicted series with Model 2. Red dotted line: predicted series with Model 3. Green - . . : predicted series with AR(4) model. Black solid line: the observed data.

for forecasting GDP and IP over horizons longer than 12 months ($h > 12$), the mixture-FAR models performed better than the AR(4) model. Combining these observations with those in the previous sections, we conclude that the mixture-FAR model performed better than the AR(4) for forecasting GPD and IP over short and long horizons.

4.4 Mixture-FAR vs nonstationary-FAR models for forecasting GDP and IP

For forecasting a nonstationary variable such as GDP and IP, a nonstationary-FAR model, wherein all the variables including the factors are nonstationary, has been proposed in the literature (Choi 2017). As in the earlier sections, we refer to this model as a nonstationary-FAR model. To implement this method, first we performed principal component analysis on XX', and chose only the nonstationary factors for use as predictors in the prediction

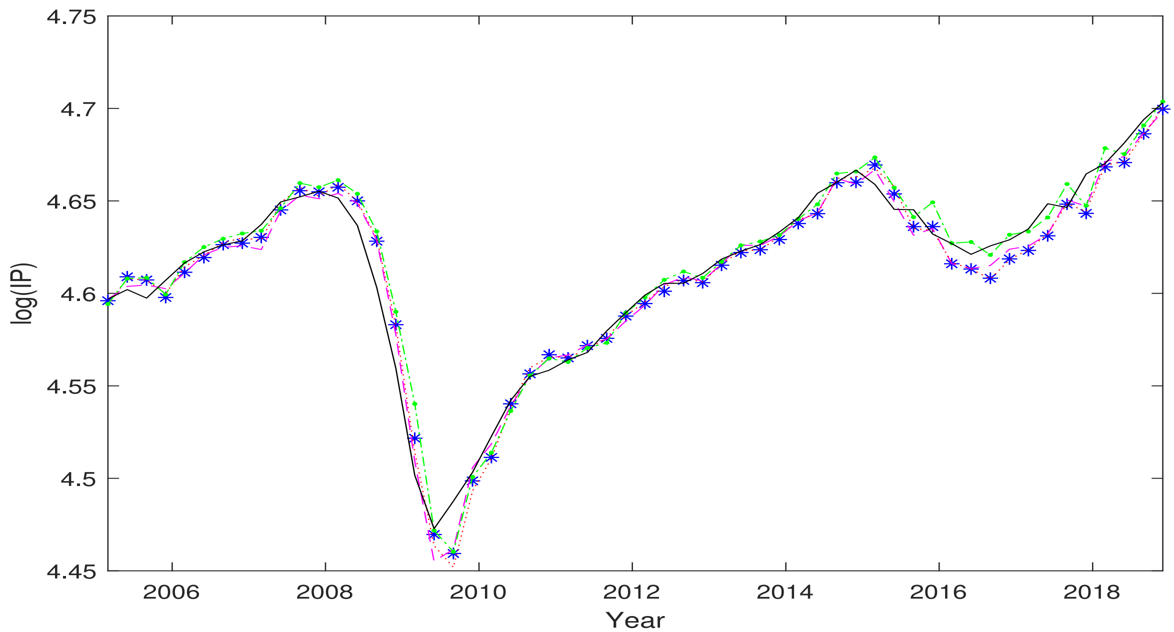

Figure 4: The observed $\log(IP)$, and plots of one-step ahead out-of-sample forecasts of $\log(IP)$ for 2006:Q1 – 2018:Q4. The other legends are the same as for Figure 3.

model (2). We wish to compare the aforementioned nonstationary-FAR method with the mixture-FAR method. To this end, we chose Model 1 as our mixture-FAR model; for the nonstationary-FAR model, we chose

$$\text{Model 5[Nonstationary-FAR]:} \quad Y_{t+h} = \beta' \tilde{E}_t + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+h}.$$

The values of R^2_{os} for these two models are provided in Table 4. Consider the entry 0.31 in the column for $\log(IP)$. This says that the *sum of squares of forecast error*[SSFE] for $\log(IP)$ over the period 2009–2018 is 31% lower for mixture-FAR compared to the nonstationary-FAR. The table also shows that the SSFE for $\log(GDP)$ over the period 2009–2018 is 50% lower for mixture-FAR compared to nonstationary-FAR. In fact, Table 4 shows that the mixture-FAR method proposed in this paper performed significantly better than the method based on a nonstationary-FAR for forecasting GDP and IP.

Table 3: Performance of mixture-FAR relative to AR(4), in terms of R^2_{os}.

	log(GDP)			log(IP)		
First estimation period	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1959:Q1 - 1999:Q4	-0.01	0.07	-0.00	-0.14	0.10	-0.21
1959:Q1 - 2005:Q4	-0.08	0.11	-0.01	0.17	0.33	0.11
1959:Q1 - 2008:Q4	0.09	0.16	0.18	0.06	0.28	-0.01

Note: Each entry in the table is the R^2_{os} for a given mixture-FAR model relative to the AR(4) model. These are for one-step ahead forecasts with expanding windows for estimation. The out-of-sample forecast period starts from the end of the estimation period and extends to the end of 2018.

4.5 Prediction intervals

We computed 95% prediction intervals, based on the asymptotic results in Section 2 and the bootstrap, for one-step-ahead prediction intervals for log(GDP) and log(IP) with expanding window for the estimation period. Assuming that the regression error ϵ_t in (2) is normally distributed, we constructed the asymptotic theory-based point-wise 95% prediction intervals for log(GDP) and log(IP) for the out-of-sample period 2006:Q1 to 2018:Q4. These intervals are shown in Figures 7 and 8 together with the observed values of log(GDP) and log(IP).

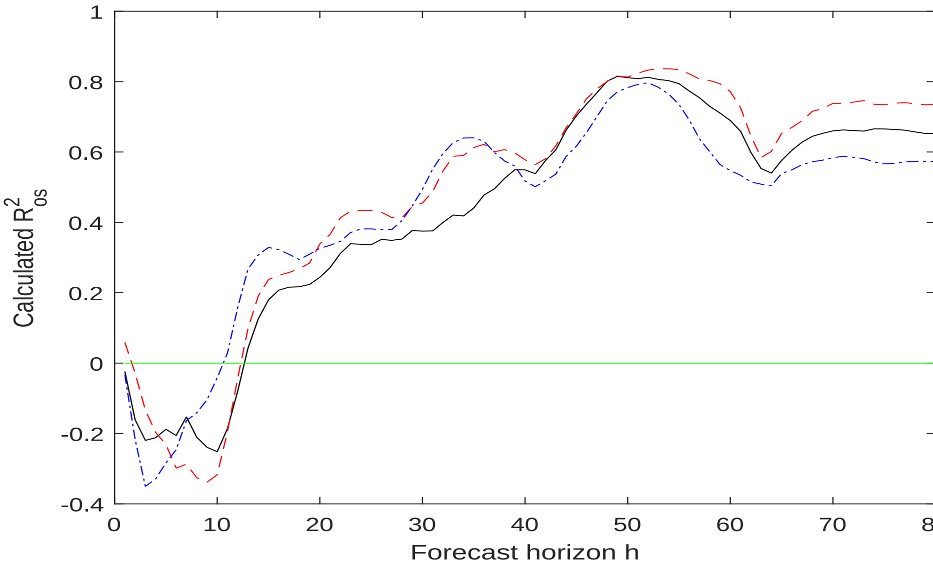

Figure 5: Performance of mixture-FAR relative to AR(4) for long-term forecast of $\log(GDP)$. The graph is a plot of R^2_{os} against the forecast horizon h. Black solid line is the R^2_{os} for the model 1 relative to AR(4). Red dash line is the R^2_{os} for model 2 relative to AR(4). Blue dotted dash line is the R^2_{os} for the model 3 relative to AR(4).

We also estimated the symmetric bootstrap t-percentile prediction intervals using residual bootstrap; to this end, we used 399 bootstrap replications. These are also shown in Figures 7 and 8. One important difference between the asymptotic theory-based and bootstrap prediction intervals is that the latter does not assume that the error distribution has a known functional form.

Figure 7 shows that, except for a very short interval around the crisis period 2009, the observed values of GDP lie within the two prediction intervals. Figure 8 shows that every observed value of IP lies within the two prediction intervals. Overall, the bootstrap prediction interval is narrower than the one based on the asymptotic distribution of the forecast, for GDP and IP. The bootstrap prediction interval for IP around the crisis period is large; this may be because the financial and economic crisis introduced large fluctuations in IP. Overall, both prediction intervals have high coverage rates for GDP and IP.

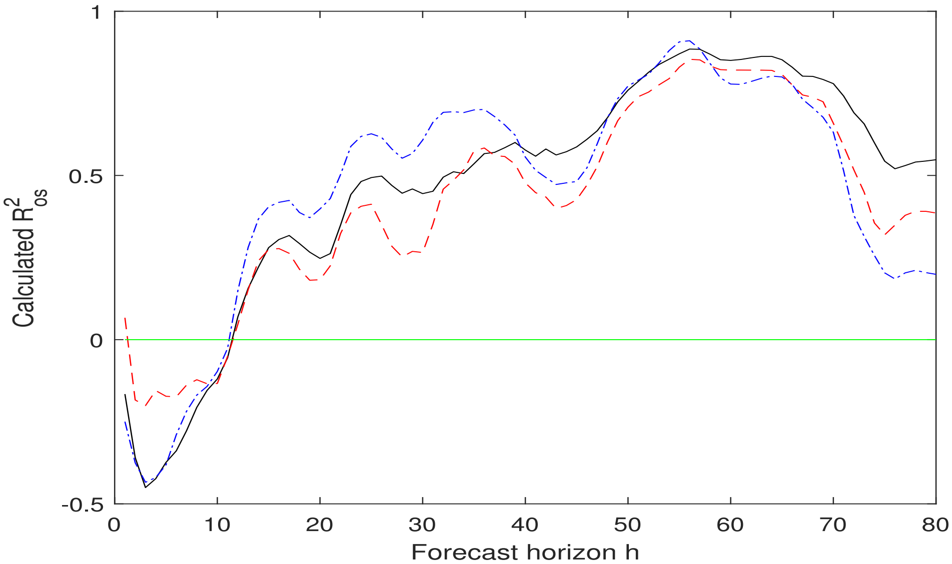

Figure 6: Performance of mixture-FAR relative to AR(4) for forecasting $\log(IP)$ over long forecast horizons. Each curve is a plot of R^2_{os} for a given mixture-FAR relative to AR(4), against the forecast horizon h. Black solid line: Model 1. Red dashed line: Model 2. Blue dotted dash line: Model 3.

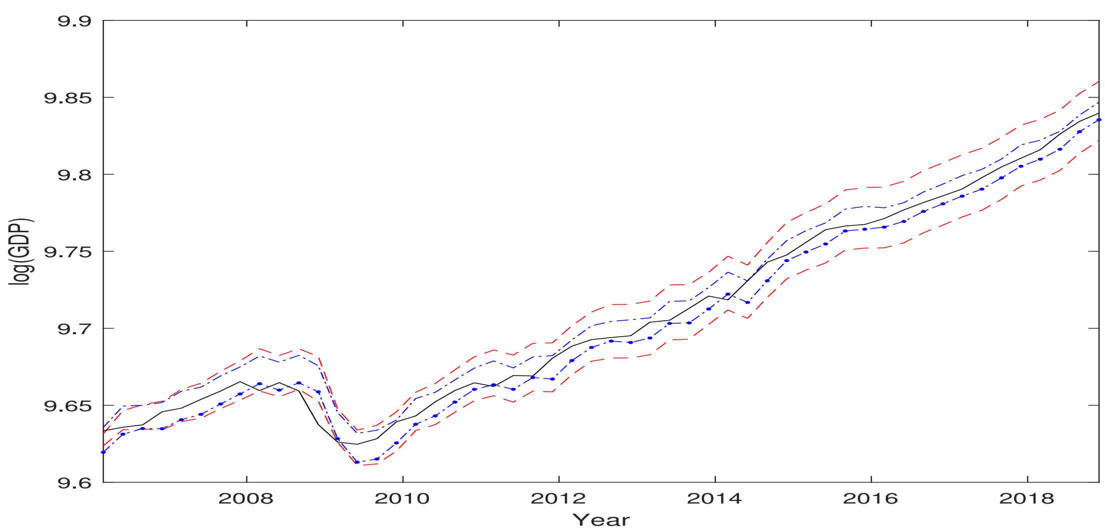

Figure 7: One-step ahead point-wise 95% prediction intervals for $\log(GDP)$ using mixture-FAR. The solid black line in the middle is a plot of the observed values of $\log(GDP)$. Red dashed line: the asymptotic theory based prediction interval. Blue dotted and dashed line: the bootstrap prediction interval.

Table 4: Values of R^2_{os} for mixture-FAR compared to the nonstationary-FAR.

First estimation period	$log(GDP)$	$log(IP)$
1959:Q1 - 1999:Q4	0.21	0.17
1959:Q1 - 2005:Q4	0.21	0.39
1959:Q1 - 2008:Q4	0.50	0.31

Note: The mixture-FAR is Model 1 and the nonstationary-FAR is Model 5.

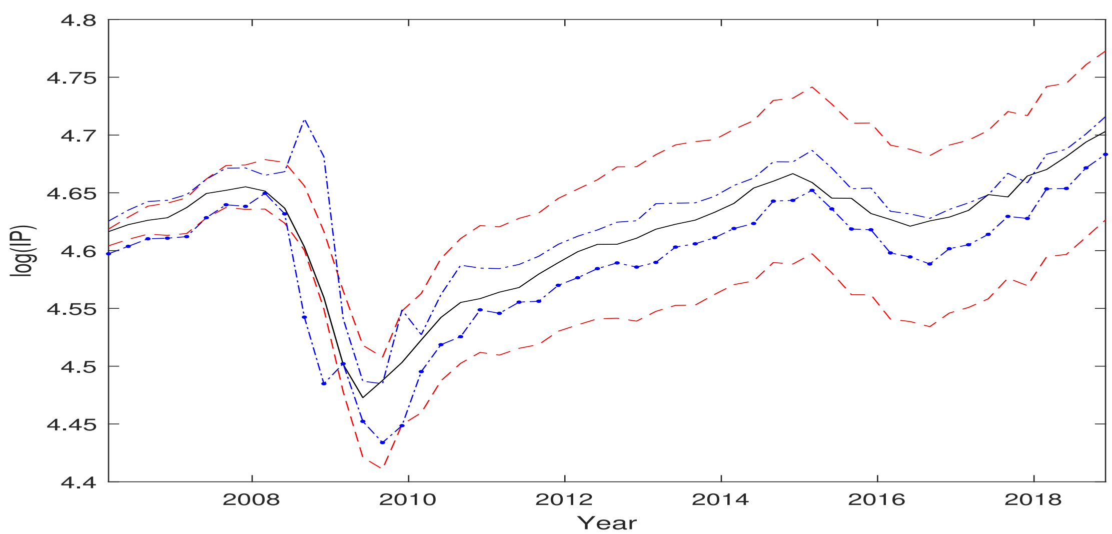

Figure 8: The legend is the same as for Figure 7, except that GDP is replaced by IP.

5 Conclusion

This paper developed methodology for forecasting nonstationary macroeconomic variables, such as GDP and industrial production[IP], when a set of panel data is available for a large number of potential predictors. We propose to estimate a small number of factors using the panel data, and use them as predictors for forecasting. The factors are chosen such that they contain a large proportion of the information in the large number of potential predictors. The validity of this method for forecasting has been established in the literature when all the variables are stationary (Bai & Ng 2006), and also when they are all nonstationary (Choi 2017), but not when they consist of a mixture of stationary and nonstationary variables. Typically, a set of panel data on such a large number of macroeconomic variables would contain mixture of stationary and nonstationay variables, which turned out be also the case in our empirical example. Therefore, the method developed in this paper is of practical significance. To use the estimated mixture of stationary and nonstationary factors as predictors, and construct an asymptotically valid prediction interval, this paper developed the methodology. In our simulation study, the mixture-FAR method developed in this paper performed significantly better than the one that uses only nonstationary variables. We applied the proposed method for forecasting GDP and IP. We assessed the out-of-sample forecast performance of the mixture-FAR relative to the corresponding nonstationary-FAR and the AR(4) models. We observed that the mixture-FAR model performed significantly better than the aforementioned two competing methods. In summary, this paper provides an improved method of forecasting a nonstationary variable using information from stationary and nonstationary variables.

References

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', Econometrica **71**(1), 135–171.

Bai, J. (2004), 'Estimating cross-section common stochastic trends in nonstationary panel data', Journal of Econometrics **122**(1), 137–183.

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', Econometrica **70**(1), 191–221.

Bai, J. & Ng, S. (2006), 'Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions', Econometrica **74**(4), 1133–1150.

Bernanke, B. S., Boivin, J. & Eliasz, P. (2005), 'Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach', The Quarterly Journal of Economics **120**(1), 387–422.

Cheng, X. & Hansen, B. E. (2015), 'Forecasting with factor-augmented regression: A frequentist model averaging approach', Journal of Econometrics **186**(2), 280–293.

Choi, I. (2017), 'Efficient estimation of nonstationary factor models', Journal of Statistical Planning and Inference **183**, 18–43.

Eickmeier, S. (2005), Common stationary and non-stationary factors in the euro area analyzed in a large-scale factor model, Discussion paper 2, Studies of the Economic Research Centre.

Gonçalves, S. & Perron, B. (2014), 'Bootstrapping factor-augmented regression models', Journal of Econometrics **182**(1), 156–173.

Ludvigson, S. C. & Ng, S. (2007), 'The empirical risk–return relation: A factor analysis approach', Journal of Financial Economics **83**(1), 171–222.

Moon, H. R. & Perron, B. (2007), 'An empirical analysis of nonstationarity in a panel of interest rates with factors', Journal of Applied Econometrics **22**(2), 383–400.

Smeekes, S. & Wijler, E. (2019), High-dimensional forecasting in the presence of unit roots and cointegration, Working paper, Dept. of Quant. Econ., Maastricht Univ.

Stock, J. H. & Watson, M. W. (1998a), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, Working Paper 6672, National Bureau of Economic Research.

Stock, J. H. & Watson, M. W. (1998b), Diffusion indexes, Working Paper 6702, National Bureau of Economic Research.

Stock, J. H. & Watson, M. W. (2002a), 'Forecasting using principal components from a large number of predictors', Journal of the American Statistical Association **97**(460), 1167–1179.

Stock, J. H. & Watson, M. W. (2002b), 'Macroeconomic forecasting using diffusion indexes', Journal of Business & Economic Statistics **20**(2), 147–162.

Stock, J. H. & Watson, M. W. (2012), 'Generalized shrinkage methods for forecasting using many predictors', Journal of Business & Economic Statistics **30**(4), 481–493.

Supplementary material for
"Time Series Forecasting using a Mixture of Stationary and Nonstationary Predictors"

Appendix A defines mathematical symbols and then recalls the necessary assumptions. Appendix B gives the proofs of Lemmas 1–3 listed in the main paper and then contains the proofs of the auxiliary lemmas used to prove the main results in Appendix C. Simulation results for bootstrap are presented in Section D.

6 Appendix A

First, we introduce the following notation and mathematical symbols:

$$\gamma_{st} = E\left(N^{-1}\textstyle\sum_{i=1}^{N} e_{is}e_{it}\right), \quad \zeta_{st} = N^{-1}\sum_{i=1}^{N}\left(e_{is}e_{it} - E(e_{is}e_{it})\right), \tag{A.1}$$

$$\eta_{st} = N^{-1}F_s'\Lambda' e_t, \qquad \xi_{st} = N^{-1}F_t'\Lambda' e_s \tag{A.2}$$

$$\tau_{ij,ts} = E(e_{it}e_{js}), \qquad \tau_{ij,t} = \tau_{ij,tt}; \tag{A.3}$$

for $i, j = 1, \ldots N; s, t = 1, \ldots, T$, and let us note that $\eta_{st} = \xi_{ts}$.

Variables:

$Y = (Y_t)_{T\times 1} = (Y_1, \ldots, Y_T)'$, uni-variate dependent variable,

$W = (W_{it})_{T\times m} = \begin{pmatrix} W_{11} & \ldots & W_{m1} \\ \vdots & \ddots & \vdots \\ W_{1T} & \ldots & W_{mT} \end{pmatrix} = (W_1', \ldots, W_T')'$; m is the number of observable (nonstationary) regressors,

$$F = (F_{it})_{T\times r} = \begin{pmatrix} F_{11} & \ldots & F_{r1} \\ \vdots & \ddots & \vdots \\ F_{1T} & \ldots & F_{rT} \end{pmatrix} = \begin{pmatrix} F_1' \\ \vdots \\ F_T' \end{pmatrix}; F' = (F_1, \ldots, F_T),$$

$$\Lambda = (\lambda_{ij})_{N\times r} = \begin{pmatrix} \lambda_{11} & \ldots & \lambda_{r1} \\ \vdots & \ddots & \vdots \\ \lambda_{1N} & \ldots & \lambda_{rN} \end{pmatrix} = \begin{pmatrix} \lambda_1' \\ \vdots \\ \lambda_N' \end{pmatrix}; \Lambda' = (\lambda_1, \ldots, \lambda_N),$$

$$e = (e_{it})_{T\times N} = \begin{pmatrix} e_{11} & \dots & e_{N1} \\ \vdots & \ddots & \vdots \\ e_{1T} & \dots & e_{NT} \end{pmatrix} = \begin{pmatrix} e_1' \\ \vdots \\ e_T' \end{pmatrix}; \; e' = (e_1, \dots, e_T).$$

Scaling matrices:

$D_{1T} = \begin{pmatrix} TI_p & 0 \\ 0 & T^{1/2}I_q \end{pmatrix}$; p, q are the number of nonstationary and stationary factors respectively.

$$D_T = \begin{pmatrix} TI_p & 0 & 0 \\ 0 & T^{1/2}I_q & 0 \\ 0 & 0 & TI_m \end{pmatrix}.$$

$$\|D_{1T}\| = O(T), \quad \|D_{1T}\|^{-1} = O(T^{-1}), \quad \left\|D_{1T}^{-1}\right\| = O(T^{-1/2}), \quad \left\|D_{1T}^{-1}\right\|^{-1} = O(T^{1/2}).$$

Matrices from Assumption 1:

$D_{1T}^{-1} \sum_{t=1}^{T} F_t F_t' D_{1T}^{-1} \xrightarrow{d} \Sigma_{F_{(r\times r)}}$; a positive definite random matrix,

$N^{-1}\Lambda'\Lambda \to \Sigma_{\Lambda_{(r\times r)}}$ or $N^{-1}\Lambda'\Lambda \xrightarrow{p} \Sigma_{\Lambda_{(r\times r)}}$; a positive definite non-random matrix.

Matrices with eigenvalues:

$\tilde{V}_{NT} = \mathrm{diag}(\tilde{Z}) \times N^{-1} D_{1T}^{-2}; \tilde{Z} = (\tilde{v}_1, \dots, \tilde{v}_r)$; the r largest eigenvalues of XX',

$\tilde{V}_{NT}^* = \mathrm{diag}(\tilde{Z}^*) \times D_{1T}^{-2}; \tilde{Z}^* = (v_1^*, \dots, v_r^*)$; the r largest eigenvalues of $N^{-1}F\Lambda'\Lambda F'$,

$V = \mathrm{diag}(Z); Z = (v_1, \dots, v_r)$; the eigenvalues of $\Sigma_\Lambda \Sigma_F$.

Rotation matrix:

$H = N^{-1}\tilde{V}_{NT}^{-1} D_{1T}^{-2} \tilde{F}' F \Lambda' \Lambda$; $r \times r$ matrix.

Parameters:

$\delta = \begin{pmatrix} \alpha' & \beta' & \omega' \end{pmatrix}' = \begin{pmatrix} \theta' & \omega' \end{pmatrix}'; \theta' = \begin{pmatrix} \alpha' & \beta' \end{pmatrix}$, where

$\alpha : p \times 1$ column vector of parameters corresponding to nonstationary factors,

$\beta : q \times 1$ column vector of parameters corresponding to stationary factors,

$\omega : m \times 1$ column vector of parameters corresponding to observable regressors.

$\hat{\delta}$ denotes the OLS estimated coefficients.

To help the reader recall the main assumptions imposed in the main paper, this appendix lists the following assumptions again. In these assumptions, $M \in \mathbb{R}$ denotes a finite constant.

Assumption 1. Factors and factor loadings

i. The strictly stationary process u_t in (4) satisfies $\max_{t\geq 1} E\|u_t\|^{4+\delta} \leqslant M$ for some $\delta > 0$.

ii. $E\|F_1\|^4 \leqslant M$ and $D_{1T}^{-1}\sum_{t=1}^{T} F_t F_t' D_{1T}^{-1} \xrightarrow{d} \Sigma_F$ as $T \to \infty$, where Σ_F is a positive definite random matrix.

iii. The number of factors r is known and does not depend on N or T. Further, factors are not cointegrated.

iv. The loadings λ_i are either deterministic and $\|\lambda_i\| \leqslant M$ satisfying $\Lambda'\Lambda/N \to \Sigma_\Lambda$ as $N \to \infty$, or they are stochastic and $E\|\lambda_i\|^4 \leqslant M$ satisfying $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda$ as $N \to \infty$, for some $r \times r$ positive definite non-random matrix Σ_Λ.

v. The eigenvalues of the matrix $\Sigma_\Lambda \Sigma_F$ are distinct, almost surely.

Assumption 2. Idiosyncratic errors

i. $E(e_{it}) = 0$ and $E|e_{it}|^8 \leqslant M$ $(i = 1, \ldots, N; t = 1, \ldots, T)$.

ii. Let $\gamma_{st} = E\left(N^{-1}\sum_{i=1}^{N} e_{is}e_{it}\right)$ $(s, t = 1, \ldots, T)$. $|\gamma_{ss}| \leqslant M (s = 1, ..., T)$, and $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} |\gamma_{st}| \leqslant M$.

iii. Let $\tau_{ij,t} = E(e_{it}e_{jt})$ $(i, j = 1, \ldots, N; t = 1, \ldots, T)$. $|\tau_{ij,t}| \leqslant |\tau_{ij}|$ for some τ_{ij} $(i, j = 1, \ldots, N;\ t = 1, \ldots, T)$, and $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} |\tau_{ij}| \leqslant M$.

iv. Let $\tau_{ij,ts} = E(e_{it}e_{js})$ $(i, j = 1, \ldots, N; t = 1, \ldots, T)$. $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |\tau_{ij,ts}| \leqslant M$.

v. $E\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{is}e_{it} - E(e_{is}e_{it})\right]\right|^4 \leqslant M$ $(t, s = 1, \ldots, T)$.

Assumption 3. Dependency among $\lambda_i, F_t,$ and e_{it}

i. $E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|D_{1T}^{-1}\sum_{t=1}^{T} F_t e_{it}\right\|^2\right) \leqslant M$, and $E(F_t\, e_{it}) = 0$ $(i = 1, \ldots, N;\ t = 1, \ldots, T)$.

ii. Let $\Gamma_t = \lim_{N\to\infty}(1/N)\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(\lambda_i\lambda_j' e_{it}e_{jt}\right)$. Then, for each fixed t, we have $N^{-1/2}\Lambda' e_t \xrightarrow{d} N(0, \Gamma_t)$ as $N \to \infty$.

iii. $E\left\|N^{-1/2}D_{1T}^{-1}\sum_{t=1}^{T}\Lambda' e_t F_t'\right\|^2 \leqslant M$.

Assumption 4. Weak dependence of idiosyncratic errors

i. $\sum_{s=1}^{T} |\gamma_{st}| \leqslant M$, $(t = 1, ..., T)$,

ii. $\sum_{j=1}^{N} |\tau_{ij}| \leqslant M$, $(i = 1, ..., N)$, where γ_{st} and τ_{ij} are defined in A.1 and Assumption 2.

Assumption 5. Weak dependence between idiosyncratic and regression error

i. $E\left|(TN)^{-1/2}\sum_{s=1}^{T-h}\sum_{i=1}^{N}\left(e_{is}e_{it}-E\left(e_{is}e_{it}\right)\right)\epsilon_{s+h}\right|^2 \leqslant M \quad (\, t=1,\ldots,T; h>0)$.

ii. $E\left\|(TN)^{-1/2}\sum_{t=1}^{T-h}\sum_{i=1}^{N}\lambda_i e_{it}\epsilon_{t+h}\right\|^2 \leqslant M$, and $E\left(\lambda_i e_{it}\epsilon_{t+h}\right) = 0 \quad (i = 1,\ldots,N; t = 1,\ldots,T)$.

Assumption 6. Moment and CLT for score vector

Let $L_t = (F_t', W_t')'$.

i. $E\left(\epsilon_{t+h}\right) = 0$ and $E\left|\epsilon_{t+h}\right|^2 < M$ $(t=1,\ldots,T)$.

ii. $D_T^{-1}\sum_{t=1}^{T} L_t L_t' D_T^{-1} \xrightarrow{d} \Sigma_L$ as $N,T\to\infty$, where Σ_L is a positive definite random matrix.

iii. $D_T^{-1}\sum_{t=1}^{T} L_t\varepsilon_{t+h} \xrightarrow{d} \Sigma_{\epsilon L}^{1/2}\times N(0,I)$, where $\Sigma_{\epsilon L}>0$ with probability one.

7 Appendix B

Before we provide the proofs of Theorems 1 and 2 in Appendix C below, we introduce a series of auxiliary lemmas.

Lemma B. 1. *Suppose that Assumptions 1-3 are satisfied. Then we have*

(i) $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{st}^2 \leqslant M$,

(ii) $T^{-1}N^{-1}\sum_{t=1}^{T}\|\Lambda' e_t\|^2 = O_P(1)$.

Proof. Part (i) Let $\rho_{st} = \gamma_{st}(\gamma_{ss}\gamma_{tt})^{-1/2}$ $(s=1,\ldots,T; t=1,\ldots,T)$. Then $|\rho_{st}| \leq 1$, and by Assumption 2(ii), we have $|\gamma_{ss}| \leqslant M$.

$$\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{st}^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_{ss}\gamma_{tt}\rho_{st}^2 \leq M\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{ss}\gamma_{tt}|^{1/2}|\rho_{st}| = M\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}| \leq M^2,$$

since $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{st}| \leqslant M$ by Assumption 2(ii).

Part (ii) Using the Assumptions 1(iv) and 2(iii), we obtain

$$\frac{1}{TN}\sum_{t=1}^{T}\|\Lambda' e_t\|^2 = \frac{1}{TN}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\lambda_i e_{it}\right)'\left(\sum_{j=1}^{N}\lambda_j e_{jt}\right) = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i'\lambda_j e_{it}e_{jt},$$

$$E\left(\frac{1}{TN}\sum_{t=1}^{T}\|\Lambda' e_t\|^2\right) = E\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i'\lambda_j e_{it}e_{jt}\right) \leqslant \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|\lambda_i'\lambda_j|\,|E(e_{it}e_{jt})|$$

$$= \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|\lambda_i'\lambda_j|\,|\tau_{ij,t}| \leqslant M^2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}| \leq M^3.$$

Therefore, $T^{-1}N^{-1}\sum_{t=1}^{T}\|\Lambda' e_t\|^2 = T^{-1}\sum_{t=1}^{T}\left\|N^{-1/2}\sum_{i=1}^{N}e_{it}\lambda_i\right\|^2 = O_P(1)$. ■

Let $\tilde{V}_{NT}^* = \text{diag}(v_1^*,\ldots,v_r^*)D_{1T}^{-2}$, where $v_1^*,\ldots,v_r^*$ are the r non-zero eigenvalues of $N^{-1}F(\Lambda'\Lambda)F'$. Similarly, let $V = \text{diag}(v_1,...,v_r)$ where $\{v_1,\ldots,v_r\}$ are the eigenvalues of $\Sigma_\Lambda\Sigma_F$.

Lemma B. 2. *Suppose that Assumptions 1-3 are satisfied. Then, as $N,T\to\infty$ with $T/N\to 0$,*

(i) $\left\|D_{1T}^{-2}\tilde{F}'\left(\frac{XX'}{N}\right)\tilde{F}D_{1T}^{-2} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| = o_P(1)$,

(ii) $N^{-1}D_{1T}^{-2}\tilde{F}'XX'\tilde{F}D_{1T}^{-2} = \tilde{V}_{NT} \xrightarrow{d} V$,

(iii) $\|H\| = \left\|\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\frac{\Lambda'\Lambda}{N}\right\| = O_P(1)$.

Proof. ***Part (i)*** Let $W = D_{1T}^{-2}\tilde{F}'\left(XX'N^{-1}\right)\tilde{F}D_{1T}^{-2} - D_{1T}^{-2}\tilde{F}'F\left(\Lambda'\Lambda N^{-1}\right)F'\tilde{F}D_{1T}^{-2}$. By substituting $X = F\Lambda' + e$, we obtain,

$$\begin{aligned}W &= N^{-1}D_{1T}^{-2}\tilde{F}'\left(F\Lambda'+e\right)\left(F\Lambda'+e\right)'\tilde{F}D_{1T}^{-2} - N^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda F'\tilde{F}D_{1T}^{-2}\\ &= N^{-1}D_{1T}^{-2}\tilde{F}'\left\{F\Lambda'\Lambda F' + F\Lambda' e' + e\Lambda F' + ee' - F\Lambda'\Lambda F'\right\}\tilde{F}D_{1T}^{-2}\\ &= N^{-1}D_{1T}^{-2}\tilde{F}'\left\{F\Lambda' e' + e\Lambda F' + ee'\right\}\tilde{F}D_{1T}^{-2} = D_{1T}^{-2}\tilde{F}'W^*,\end{aligned}$$

where $W^* = N^{-1}\{F\Lambda'e' + e\Lambda F' + ee'\}\tilde{F}D_{1T}^{-2}$. Using Cauchy Schwarz inequality, we have $\|W\|^2 \leqslant \left\|D_{1T}^{-2}\tilde{F}\right\|^2\|W^*\|^2$. Consider W^*. Let $W_t^* = N^{-1}D_{1T}^{-2}\tilde{F}'\{ee_t + F\Lambda'e_t + e\Lambda F_t\}$ for $t = 1,\ldots,T$. Then,

$$\begin{aligned}\|W^*\|^2 = \left\|(W^*)'\right\|^2 &= \left\|\frac{1}{N}D_{1T}^{-2}\tilde{F}'\left(e\Lambda F' + F\Lambda' e' + ee'\right)\right\|^2\\ &= \left\|\frac{1}{N}D_{1T}^{-2}\tilde{F}'\left(e\Lambda\left(F_1,\ldots,F_T\right) + F\Lambda'\left(e_1,\ldots,e_T\right) + e\left(e_1,\ldots,e_T\right)\right)\right\|^2\\ &= \|(W_1^*,\ldots,W_T^*)\|^2 = \sum_{t=1}^{T}\|W_t^*\|^2,\end{aligned}$$

where $W_t^* = N^{-1}D_{1T}^{-2}\tilde{F}'(e\Lambda F_t+F\Lambda' e_t+ee_t)$. From the fact that $(a+b+c+d)^2 \leqslant 4\left(a^2+b^2+c^2+d^2\right)$, we obtain,

$$
\begin{aligned}
\|W^*\|^2 &= \sum_{t=1}^{T}\left\|N^{-1}D_{1T}^{-2}\tilde{F}'\left\{ee_t+F\Lambda' e_t+e\Lambda F_t\right\}\right\|^2 \\
&= \sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}+D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}+D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}+D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2 \\
&\leqslant 4\sum_{t=1}^{T}\left(\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2+\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2+\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2+\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2\right) \\
&= 4\left(\sum_{t=1}^{T}a_t+\sum_{t=1}^{T}b_t+\sum_{t=1}^{T}c_t+\sum_{t=1}^{T}d_t\right) \qquad (say),
\end{aligned}
$$

where

$$
a_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2, b_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2, c_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2, d_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2. \tag{B.1}
$$

Now, consider each term separately. By Cauchy Schwarz inequality,

$$
\frac{1}{T}\sum_{t=1}^{T}a_t = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2 \leqslant \|D_{1T}^{-1}\|^2\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{st}^2
$$

From the normalization condition $D_{1T}^{-2}\tilde{F}'\tilde{F} = I_r$, we have $\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2 = O_P(1)$. Using Lemma (B.1), we have $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{st}^2 = O(1)$. Therefore, we have $T^{-1}\sum_{s=1}^{T}a_t = O_P\left(\|D_{1T}^{-1}\|^2\right)$.
Since $\left\|D_{1T}^{-1}\right\|^2 = O(T^{-1})$, we obtain

$$
\sum_{t=1}^{T}a_t = O_P\left(T\|D_{1T}^{-1}\|^2\right) = O_P(1). \tag{B.2}
$$

Now, consider the the second term. Using the Cauchy Schwarz inequality, we have

$$
\frac{1}{T}\sum_{t=1}^{T}b_t = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2 \leqslant \|D_{1T}^{-1}\|^2\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\underbrace{\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\zeta_{st}|^2}_{K_1 \text{ (say)}}.
$$

Using Assumption 2 (v), we have $E\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{is}e_{it}-E\left(e_{is}e_{it}\right)\right]\right|^4 = N^2E\left|\zeta_{st}\right|^4 \leqslant M < \infty$. Therefore,

$$
E\left|K_1\right| \leqslant \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\zeta_{st}\right|^2 \leqslant \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{M_1}{N} = O\left(\frac{T}{N}\right),
$$

where M_1 is a finite constant. Thus, together with the fact that, $T||D_{1T}^{-2}|| = O(1)$, we obtain

$$\frac{1}{T}\sum_{t=1}^{T} b_t = O_P\left(\left\|D_{1T}^{-1}\right\|^2\right) O_P\left(\frac{T}{N}\right) = O_P\left(\frac{1}{N}\right).$$

Hence, we have

$$\sum_{t=1}^{T} b_t = O_P\left(\frac{T}{N}\right). \tag{B.3}$$

Now, consider the third term. Using Cauchy Schwartz inequality, Assumption 1, normalization condition on factors, and Lemma (B.1), we obtain,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} c_t &= \frac{1}{T}\textstyle\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\frac{F_s'\Lambda' e_t}{N}\right\|^2 \\ &\leqslant \frac{1}{N}\underbrace{\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\Lambda' e_t}{\sqrt{N}}\right\|^2}_{O_P(1)}\underbrace{\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s F_s'\right\|^2}_{K_2\ \text{(say)}},\end{aligned} \tag{B.4}$$

where we have

$$K_2 = O_P\left(\textstyle\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\right) O_P\left(\sum_{s=1}^{T}\left\|D_{1T}^{-1}F_s\right\|^2\right) = O_P(1)$$

Therefore, $T^{-1}\sum_{t=1}^{T} c_t = O_P\left(N^{-1}\right)$. Hence we have,

$$\sum_{t=1}^{T} c_t = O_P\left(\frac{T}{N}\right). \tag{B.5}$$

Similarly, we can show that $\sum_{t=1}^{T} d_t = O_P\left(TN^{-1}\right)$. Using Cauchy Schwarz inequality,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} d_t &= \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\xi_{st}\tilde{F}_s\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\frac{F_t'\Lambda' e_s}{N}\tilde{F}_s\right\|^2 \\ &\leqslant \frac{1}{TN}\left(\sum_{t=1}^{T}\left\|D_{1T}^{-1}F_t\right\|^2\right)\left(\frac{1}{N}\sum_{s=1}^{T}\left\|\Lambda' e_s\right\|^2\right)\left(\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\right) = O_P\left(\frac{1}{N}\right).\end{aligned}$$

Thus,

$$\sum_{t=1}^{T} d_t = O_P\left(\frac{T}{N}\right). \tag{B.6}$$

From equations (B.2)–(B.6), we have

$$\|W^*\|^2 \leqslant 4\left(\sum_{t=1}^{T} a_t + \sum_{t=1}^{T} b_t + \sum_{t=1}^{T} c_t + \sum_{t=1}^{T} d_t\right) = O_P(1) + O_P\left(\frac{T}{N}\right).$$

Thus,

$$\|W\| \leqslant \left\|D_{1T}^{-1}\right\| \left\|D_{1T}^{-1}\tilde{F}'\right\| \|W^*\| = O\left(\left\|D_{1T}^{-1}\right\|\right) O_P(1) \left\{O_P(1) + O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right)\right\} = o_P(1),$$

as $N, T \to \infty$.

Part (ii) By the definition of eigenvalues and eigenvectors, $N^{-1}XX'\tilde{F}D_{1T}^{-2} = \tilde{F}\tilde{V}_{NT}$, and the normalization condition, $D_{1T}^{-2}\tilde{F}'\tilde{F} = I_r$, we have, $N^{-1}D_{1T}^{-2}\tilde{F}'XX'\tilde{F}D_{1T}^{-2} = \tilde{V}_{NT}$. Now, together with part (i), we can write

$$\left\|\tilde{V}_{NT} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| = o_P(1). \tag{B.7}$$

Furthermore, recall that $\tilde{V}_{NT}^*$ is the diagonal matrix with largest r eigenvalues of $F\left(\Lambda'\Lambda N^{-1}\right)F'$ multiplied by D_{1T}^{-2} and $\tilde{F}^*$ be the corresponding eigenvector matrix such that $D_{1T}^{-2}\tilde{F}^{*\prime}\tilde{F}^* = I_r$. Then, using similar arguments as in *part (i)*, we have

$$\left\|D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2} - D_{1T}^{-2}\tilde{F}^{*\prime}F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}^*D_{1T}^{-2}\right\| = o_P(1). \tag{B.8}$$

Again, by the definition of eigenvalues and eigenvectors, and the normailzation condition, we may write $N^{-1}D_{1T}^{-2}\tilde{F}^{*\prime}(F\Lambda'\Lambda F')\tilde{F}^*D_{1T}^{-2} = \tilde{V}_{NT}^*$. Then, equation (B.8) gives

$$\left\|\tilde{V}_{NT}^* - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| = o_P(1). \tag{B.9}$$

Therefore, we have,

$$\begin{aligned}
\left\|\tilde{V}_{NT} - \tilde{V}_{NT}^*\right\| &= \left\|\tilde{V}_{NT} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2} - \left(\tilde{V}_{NT}^* - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right)\right\| \\
&\leqslant \left\|\tilde{V}_{NT} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| + \left\|\tilde{V}_{NT}^* - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| \\
&= o_P(1).
\end{aligned}$$

Thus, we have $\tilde{V}_{NT} = \tilde{V}_{NT}^* + o_P(1)$.

Since the r largest eigenvalues of $F\left(\Lambda'\Lambda N^{-1}\right)F'$ are the same as those of $\left(\Lambda'\Lambda N^{-1}\right)F'F$, elements of $\tilde{V}_{NT}^*$ are equal to the eigenvalues of $(\Lambda'\Lambda N^{-1})F'F$ multiplied by D_{1T}^{-2}. Then, under the Assumption 1, we have that $\tilde{V}_{NT}^*$ converges, in distribution, to V, positive definite diagonal matrix with eigenvalues of $\Sigma_\Lambda\Sigma_F$. Therefore, $\tilde{V}_{NT} \xrightarrow{d} V$ and $\tilde{V}_{NT}^{-1} = O_P(1)$.

Part (iii) This part directly holds form the *Part (ii)*, and Assumption 1.

$$\|H\| = \left\|\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\frac{\Lambda'\Lambda}{N}\right\| \leq \left\|\tilde{V}_{NT}^{-1}\right\| \left\|D_{1T}^{-2}\tilde{F}'F\right\| \left\|\frac{\Lambda'\Lambda}{N}\right\| = O_P(1).$$

■

The next lemma proves the consistency of factors, and is stated in the main paper.

Lemma 1. *Suppose that Assumptions 1– 3 are satisfied. Let* $\delta_{NT}^{-1} = \max\left[N^{-1/2}, ||D_{1T}^{-1}||\right]$. *Then, there exists an* $(r \times r)$ *non-singular matrix* H*, called a rotation matrix, such that*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2 = O_P\left(\delta_{NT}^{-2}\right).$$

Proof of Lemma 1. By the identity $\tilde{F} = N^{-1}X\tilde{\Lambda}\tilde{V}_{NT}^{-1}$, as stated in Bai and Ng(2002), and $\tilde{\Lambda} = X'\tilde{F}D_{1T}^{-2}$, we have $\tilde{F} = N^{-1}X(X'\tilde{F}D_{1T}^{-2})\tilde{V}_{NT}^{-1}$. Using $H = \tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\left\{N^{-1}\Lambda'\Lambda\right\}$ and expanding XX', we obtain

$$\begin{aligned}
\tilde{F} - FH' &= \frac{1}{N}XX'\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} - F\left(\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\frac{\Lambda'\Lambda}{N}\right)' \\
&= \frac{1}{N}XX'\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} - F\frac{\Lambda'\Lambda}{N}F'\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} \\
&= \left\{\frac{1}{N}XX' - \frac{F\Lambda'\Lambda F'}{N}\right\}\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} \\
&= \frac{1}{N}\left\{F\Lambda'e' + e\Lambda F' + ee'\right\}\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1}.
\end{aligned}$$

Since $\tilde{F} = \left(\tilde{F}_1, \ldots, \tilde{F}_T\right)'$, and $F = (F_1, \ldots, F_T)'$, in vector notation, the above equation becomes

$$\begin{aligned}
\tilde{F}_t - HF_t &= \tilde{V}_{NT}^{-1}D_{1T}^{-2}\frac{1}{N}\tilde{F}'\left\{ee_t + F\Lambda'e_t + e\Lambda F_t\right\} \\
&= \tilde{V}_{NT}^{-1}D_{1T}^{-2}\left\{\frac{1}{N}\sum_{s=1}^{T}\tilde{F}_s e_s' e_t + \frac{1}{N}\sum_{s=1}^{T}\tilde{F}_s F_s'\Lambda' e_t + \frac{1}{N}\sum_{s=1}^{T}\tilde{F}_s e_s'\Lambda F_t\right\}.
\end{aligned}$$

Then we have,

$$\tilde{F}_t - HF_t = \tilde{V}_{NT}^{-1}\left\{D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\}, \quad \text{(B.10)}$$

where $\gamma_{st}, \zeta_{st}, \eta_{st}$ and ξ_{st} as defined before.

Since $\tilde{V}_{NT}^{-1} = O_P(1)$ by Lemma B.2, and the fact that $(a+b+c+d)^2 \leqslant 4\left(a^2+b^2+c^2+d^2\right)$, we obtain $\left\|\tilde{F}_t - HF_t\right\|^2 \leq 4\left(a_t + b_t + c_t + d_t\right)$ where,

$$a_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2, b_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2, c_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2, d_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2.$$

Then, $T^{-1}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2 \leqslant T^{-1}\sum_{t=1}^{T}4\left(a_t + b_t + c_t + d_t\right)$. In the proof of Lemma (B.2) (equations (B.2)–(B.6)), we have

$$\frac{1}{T}\textstyle\sum_{t=1}^{T}a_t = O_P\left(\left\|D_{1T}^{-2}\right\|\right), \quad \displaystyle\frac{1}{T}\sum_{t=1}^{T}b_t = O_P\left(\frac{1}{N}\right), \quad \text{(B.11)}$$

$$\frac{1}{T}\textstyle\sum_{t=1}^{T}c_t = O_P\left(\frac{1}{N}\right), \quad \displaystyle\frac{1}{T}\sum_{t=1}^{T}d_t = O_P\left(\frac{1}{N}\right). \quad \text{(B.12)}$$

Thus, we obtain

$$\frac{1}{T}\sum_{t=1}^{T}\left|\left|\tilde{F}_t - HF_t\right|\right|^2 = \left\{O_P\left(||D_{1T}^{-2}||\right) + O_P\left(N^{-1}\right)\right\} = \max[O_P\left(||D_{1T}^{-2}||\right), O_P\left(N^{-1}\right)] = O_P\left(\delta_{NT}^{-2}\right),$$

where $\delta_{NT}^{-1} = \max\left[N^{-1/2}, ||D_{1T}^{-1}||\right]$. ■

Lemma B. 3. *Suppose that Assumptions 1-4 satisfied. Then, as $N, T \to \infty$, we have*

(i) $A_{1t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right)$

(ii) $A_{2t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right)$,

(iii) $A_{3t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} = O_P\left(N^{-1/2}\right)$,

(iv) $A_{4t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right)$.

Proof. ***Part (i)*** Let us write

$$A_{1t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} = D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}HF_s\gamma_{st} = A_{11t} + HA_{12t}, \quad \text{(B.13)}$$

and consider each part separately. Using Cauchy Schwartz inequality, Assumption 4, Lemma (B.1), and Lemma (1), and the fact that $\left\|D_{1T}^{-2}\right\| = O(T^{-1})$, we have,

$$\begin{aligned}\|A_{11t}\| &= \left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}\right\| \\ &\leqslant \left\|D_{1T}^{-2}\right\| \underbrace{\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}}_{O_P\left(\sqrt{T}\delta_{NT}^{-1}\right)} \underbrace{\left(\sum_{s=1}^{T}|\gamma_{st}|^2\right)^{1/2}}_{O(1)} = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right),\end{aligned}$$

$$\begin{aligned}E\|A_{12t}\| &= E\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s\gamma_{st}\right\| \leqslant \left\|D_{1T}^{-2}\right\|\sum_{s=1}^{T}E\left\|F_s\gamma_{st}\right\| \leqslant ||D_{1T}^{-2}||\sum_{s=1}^{T}\left(E||F_s||^2\right)^{1/2}\left(E\left|\gamma_{st}\right|^2\right)^{1/2} \\ &\leqslant M||D_{1T}^{-2}||\sum_{s=1}^{T}|\gamma_{st}| = O(||D_{1T}^{-2}||),\end{aligned}$$

since we assume that $\left(E||F_t||^4\right) \leqslant M$ and $\sum_{s=1}^{T}|\gamma_{st}| \leqslant M$ for some finite constant M in the Assumptions 1 and 4(i). Then, together with $\|H\| = O_P(1)$ from Lemma (B.2), and the fact that $\delta_{NT}^{-1} = \max[N^{-1/2}, ||D_{1T}^{-1}||]$, we obtain

$$A_{1t} = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right) + O_P(1)O_P\left(||D_{1T}^{-2}||\right) = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right).$$

Part (ii) Similar to Part (i), we can decompose A_{2t} as follows:

$$A_{2t} = D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} = D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \zeta_{st} + D_{1T}^{-2} \sum_{s=1}^{T} HF_s \zeta_{st} = A_{21t} + HA_{22t}. \tag{B.14}$$

By Cauchy Schwarz inequality and Assumption 2(v), we have

$$\|A_{21t}\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \zeta_{st} \right\| \leqslant \|D_{1T}^{-2}\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \underbrace{\left(\sum_{s=1}^{T} \|\zeta_{st}\|^2 \right)^{1/2}}_{A_{23t}},$$

where we have

$$\begin{aligned} E(A_{23t}^2) &= E\left(\sum_{s=1}^{T} \left\| \frac{1}{N} \sum_{i=1}^{N} (e_{is}e_{it} - E(e_{is}e_{it})) \right\|^2 \right) \\ &\leq \frac{1}{N} \left(\sum_{s=1}^{T} E \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (e_{is}e_{it} - E(e_{is}e_{it})) \right|^2 \right) = O\left(\frac{T}{N} \right). \end{aligned}$$

Then, together with $T||D_{1T}^{-2}|| = O(1)$, we get

$$A_{21t} = O\left(||D_{1T}^{-2}||\right) O_P\left(T^{1/2}\delta_{NT}^{-1}\right) O_P\left(T^{1/2}N^{-1/2}\right) = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right).$$

Consider the second term of A_{2t}. Again by Cauchy Schwarz inequality and $T||D_{1T}^{-2}|| = O(1)$,

$$\begin{aligned} E\|A_{22t}\| &= E\left\| D_{1T}^{-2} \sum_{s=1}^{T} F_s \zeta_{st} \right\| \leqslant \|D_{1T}^{-2}\| \sum_{s=1}^{T} \left(E\|F_s\|^2 \right)^{1/2} \left(E|\zeta_{st}|^2 \right)^{1/2} \\ &\leqslant M \|D_{1T}^{-2}\| \sum_{s=1}^{T} \left(E|\zeta_{st}|^2 \right)^{1/2} = O\left(\frac{||D_{1T}^{-1}||}{\sqrt{N}} \right), \end{aligned}$$

thus we have

$$A_{22t} = O_P\left(\frac{||D_{1T}^{-1}||}{\sqrt{N}} \right),$$

where the second inequality holds as we assume $E\|F_s\|^4 \leqslant M < \infty$ in Assumption 1(ii), and the second equality holds by following the Assumption 2(v).

Then, together with Lemma (B.2), $\|H\| = O_P(1)$, we obtain

$$A_{2t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right) = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right).$$

Part (iii) Again, as in the Part (i), let us write,

$$A_{3t} = D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} = D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \eta_{st} + HD_{1T}^{-2} \sum_{s=1}^{T} F_s \eta_{st} = A_{31t} + HA_{32t}.$$

By Cauchy Schwarz inequality, Lemma (1), Lemma (B.1), and Assumptions 1 and 3, we have,

$$\|A_{31t}\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \eta_{st} \right\| \leqslant \|D_{1T}^{-2}\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\underbrace{\sum_{s=1}^{T} \|\eta_{st}\|^2}_{A_{33t}} \right)^{1/2},$$

where we have

$$E\left(A_{33t}\right) = E\left(\sum_{s=1}^{T} \left\| \frac{F_s'\Lambda' e_t}{N} \right\|^2 \right) \leqslant \sum_{s=1}^{T} E \left\| \frac{F_s'\Lambda' e_t}{N} \right\|^2$$
$$\leqslant \sum_{s=1}^{T} E\left\|F_s'\right\|^2 E\left\| \frac{\Lambda' e_t}{N} \right\|^2 = E\left\| \frac{\Lambda' e_t}{N} \right\|^2 \sum_{s=1}^{T} E\left\|F_s'\right\|^2 = O\left(\frac{T}{N}\right).$$

Thus, we have $A_{31t} = O\left(\|D_{1T}^{-2}\|\right) O_P\left(\sqrt{T}\delta_{NT}^{-1}\right) O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right) = O_P\left(\frac{1}{\sqrt{N}\delta_{NT}}\right)$.

Consider the second part of A_{3t},

$$\|A_{32t}\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} F_s \eta_{st} \right\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} F_s F_s' \frac{\Lambda' e_t}{N} \right\|$$
$$\leqslant \left\| \frac{\Lambda' e_t}{N} \right\| \left\| \sum_{s=1}^{T} D_{1T}^{-1} F_s F_s' D_{1T}^{-1} \right\| \leqslant \left\| \frac{\Lambda' e_t}{N} \right\| \left(\sum_{s=1}^{T} \|D_{1T}^{-1} F_s\|^2 \right) = O_P\left(N^{-1/2}\right).$$

Therefore, we have $A_{3t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right) + O_P(1) O_P\left(N^{-1/2}\right) = O_P\left(N^{-1/2}\right)$.

Part (iv) We can similarly decompose A_{4t} as follows:

$$A_{4t} = D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} = D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \xi_{st} + D_{1T}^{-2} \sum_{s=1}^{T} HF_s \xi_{st} = A_{41t} + HA_{42t}.$$

Using Cauchy Schwartz inequality, Lemma (1), Lemma (B.1), and Assumptions 1 and 3, we obtain,

$$\|A_{41t}\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \xi_{st} \right\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) F_t' \frac{\Lambda' e_s}{N} \right\|$$
$$\leqslant \|D_{1T}^{-2}\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\underbrace{\sum_{s=1}^{T} \left\| F_t' \frac{\Lambda' e_s}{N} \right\|^2}_{A_{43t}} \right)^{1/2},$$

where we have

$$E(A_{43t}) = E\left(\sum_{s=1}^{T} \left\| F_t' \frac{\Lambda' e_s}{N} \right\|^2 \right) \leqslant \sum_{s=1}^{T} E \left\| F_t' \frac{\Lambda' e_s}{N} \right\|^2 \leqslant \sum_{s=1}^{T} E\left\|F_t\right\|^2 E\left\| \frac{\Lambda' e_s}{N} \right\|^2$$
$$= E\left\|F_t\right\|^2 \sum_{s=1}^{T} E\left\| \frac{\Lambda' e_s}{N} \right\|^2 = O\left(\frac{T}{N}\right),$$

and thus

$$\|A_{41t}\| = \left\|D_{1T}^{-2}\right\| O_P\left(\frac{\sqrt{T}}{\delta_{NT}}\right) O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right) = O_P\left(\frac{1}{\sqrt{N}\delta_{NT}}\right).$$

Consider the second part of A_{4t}. Using Cauchy Schwartz inequality, Assumptions 1 and 3(iii), we have,

$$\begin{aligned} E\|A_{42t}\| &= E\left\|D_{1T}^{-2}\sum_{s=1}^{T}\frac{F_t'\Lambda' e_s F_s'}{N}\right\| \leqslant \left\|D_{1T}^{-1}\right\| E\left\|D_{1T}^{-1}F_t'\sum_{s=1}^{T}\frac{\Lambda' e_s F_s'}{N}\right\| \\ &\leqslant \left\|D_{1T}^{-1}\right\|\left(E\|F_t\|^2\right)^{1/2}\left(E\left\|D_{1T}^{-1}\sum_{s=1}^{T}\frac{\Lambda' e_s F_s'}{N}\right\|^2\right)^{1/2} = O\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right). \end{aligned}$$

Thus, we have $||A_{4t}|| = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(||D_{1T}^{-1}||N^{-1/2}\right) = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right).$ ■

Lemma 2. *Suppose that Assumptions 1-4 are satisfied. Then,* $D_{1T}^{-2}\tilde{F}'F \xrightarrow{d} Q$ *as* $N,T\to\infty$, *where* $Q = V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2}$ *is a random matrix,* $V = \mathrm{diag}(v_1,...,v_r)$ *with* $\{v_1,\dots,v_r\}$ *denoting the eigenvalues of* $\Sigma_\Lambda\Sigma_F$, *and* Υ *is the corresponding matrix formed by scaled eigenvectors such that* $\Upsilon'\Upsilon = I_r$.

Proof. Since $\tilde{V}_{NT}$ is the diagonal matrix of r largest eigenvalues of XX' multiplied by $N^{-1}D_{1T}^{-2}$, we have $XX'\tilde{F}N^{-1}D_{1T}^{-2} = \tilde{F}\tilde{V}_{NT}$. Now, multiplying both side by $\left(\Lambda'\Lambda N^{-1}\right)^{1/2}\left(D_{1T}^{-2}F'\right)$, we have

$$\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(D_{1T}^{-2}F'\right)XX'\tilde{F}\frac{1}{N}D_{1T}^{-2} = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(D_{1T}^{-2}F'\right)\tilde{F}\tilde{V}_{NT}. \tag{B.15}$$

Since $X = F\Lambda' + e$, we can expand XX', and we obtain,

$$XX'\tilde{F}\frac{1}{N}D_{1T}^{-2} = \frac{F\Lambda'\Lambda F'}{N}\tilde{F}D_{1T}^{-2} + \left(\frac{F\Lambda' e' + e\Lambda F' + ee'}{N}\right)\tilde{F}D_{1T}^{-2}.$$

Then, we may rewrite equation (B.15) as

$$\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'\left(\frac{F\Lambda'\Lambda F'}{N}\tilde{F}D_{1T}^{-2} + \left(\frac{F\Lambda' e' + e\Lambda F' + ee'}{N}\right)\tilde{F}D_{1T}^{-2}\right) = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'\tilde{F}\tilde{V}_{NT}$$

$$\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'F\left(\frac{\Lambda'\Lambda}{N}\right)\left(F'\tilde{F}D_{1T}^{-2}\right) + A_{NT} = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'\tilde{F}\tilde{V}_{NT},$$

where $A_{NT} = \left(N^{-1}\Lambda'\Lambda\right)^{1/2} D_{1T}^{-2}F'\left(F\Lambda' e' + e\Lambda F' + ee'\right)\tilde{F}N^{-1}D_{1T}^{-2}$. We may show that $A_{NT} = o_P(1)$.

Using the proof of Lemma (1), we have $N^{-1}\left(F\Lambda' e' + e\Lambda F' + ee'\right)\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} = \tilde{F} - FH'$.

Therefore, we can write

$$\begin{aligned}
A_{NT} &= \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2} F' \left(F\Lambda' e' + e\Lambda F' + ee'\right) \tilde{F} \frac{1}{N} D_{1T}^{-2} \\
&= \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} \left(D_{1T}^{-2} F' \left(\tilde{F} - FH'\right)\right) \tilde{V}_{NT} \\
&= \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} \left(D_{1T}^{-2} \sum_{t=1}^{T} F_t \left(\tilde{F}_t - HF_t\right)'\right) \tilde{V}_{NT}.
\end{aligned}$$

Using Cauchy-Schwarz inequality and Lemma (B.2), $\tilde{V}_{NT} = O_P(1)$, we have

$$\begin{aligned}
\|A_{NT}\|^2 &= \left\| \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} \left(D_{1T}^{-2} \sum_{t=1}^{T} F_t \left(\tilde{F} - HF_t\right)'\right) \tilde{V}_{NT} \right\|^2 \\
&\leqslant \left\|\frac{\Lambda'\Lambda}{N}\right\| \left\| \sum_{t=1}^{T} D_{1T}^{-2} F_t \left(\tilde{F}_t - HF_t\right)' \right\|^2 \left\|\tilde{V}_{NT}\right\|^2 \\
&\leqslant \left\|\frac{\Lambda'\Lambda}{N}\right\| \left\|D_{1T}^{-1}\right\|^2 \left(\sum_{t=1}^{T} \left\|D_{1T}^{-1} F_t\right\|^2\right) \left(\sum_{t=1}^{T} \left\|\tilde{F}_t - HF_t\right\|^2\right)
\end{aligned}$$

where Lemma (B.2) has been used for the last inequality.

Using Assumption 1, we have $\left\|N^{-1}\Lambda'\Lambda\right\| = O_P(1)$ and $\left(\sum_{t=1}^{T} \left\|D_{1T}^{-1} F_t\right\|^2\right) = O_P(1)$. Also, by Lemma (1), $\left(T^{-1}\sum_{t=1}^{T} \left\|\tilde{F}_t - HF_t\right\|^2\right) = O_P(\delta_{NT}^{-2})$. Then, as $T\left\|D_{1T}^{-2}\right\| = O(1)$, we obtain $A_{NT} = O_P\left(\delta_{NT}^{-1}\right) = o_P(1)$.

Therefore, we have

$$\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} \left(D_{1T}^{-2} F'F\right) \left(\frac{\Lambda'\Lambda}{N}\right) \left(F'\tilde{F} D_{1T}^{-2}\right) + o_P(1) = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} \left(D_{1T}^{-2} F'\tilde{F}\right) \tilde{V}_{NT}. \tag{B.16}$$

Let $B_{NT} = \left(N^{-1}\Lambda'\Lambda\right)^{1/2} \left(D_{1T}^{-2} F'F\right) \left(N^{-1}\Lambda'\Lambda\right)^{1/2}$ and

$$C_{NT} = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} \left(F'\tilde{F} D_{1T}^{-2}\right). \tag{B.17}$$

We may show that, $D_{1T}^{-2} F'\tilde{F} = F'\tilde{F} D_{1T}^{-2}$, asymptotically. Consider,

$$D_{1T}^{-2} F'\tilde{F} = D_{1T}^{-2} F' \left(\tilde{F} - FH' + FH'\right) = D_{1T}^{-2} F'FH' + D_{1T}^{-2} F' \left(\tilde{F} - FH'\right).$$

Using Cauchy Schwartz inequality, Lemma (1) and Assumption 1, we have,

$$\left\|D_{1T}^{-2} F' \left(\tilde{F} - FH'\right)\right\| \leqslant \left\|D_{1T}^{-1}\right\| \left\|D_{1T}^{-1} F'\right\| \left\|\tilde{F} - FH'\right\| = o_P(1).$$

Thus, we have $D_{1T}^{-2} F'\tilde{F} = D_{1T}^{-2}(F'F)H' + o_P(1)$.

Since we assume that the factors are not cointegrated, and define the rotation matrix, H, to be asymptotically diagonal (with ± 1), we have $D_{1T}^{-2}F'FH'$ as a block diagonal matrix. This implies that $F'\tilde{F}$ is asymptotically block diagonal. Therefore, we can write $D_{1T}^{-2}F'\tilde{F} = F'\tilde{F}D_{1T}^{-2}$ asymptotically.

Then, we can rewrite the equation (B.16) as

$$(B_{NT}C_{NT} + A_{NT}) = C_{NT}\tilde{V}_{NT} \quad \text{or} \quad \left(B_{NT} + A_{NT}C_{NT}^{-1}\right)C_{NT} = C_{NT}\tilde{V}_{NT}.$$

Since $\tilde{V}_{NT}$ is diagonal, it follows that the columns of C_{NT} are eigenvectors of the matrix $B_{NT} + A_{NT}C_{NT}^{-1}$. However, this C_{NT} is not of unit length. Let $\Upsilon_{NT} = C_{NT}\tilde{V}_{NT}^{\dagger -1/2}$, where $\tilde{V}_{NT}^{\dagger}$ is a diagonal matrix with diag $\left(\tilde{V}_{NT}^{\dagger}\right) = \text{diag}\left(C_{NT}'C_{NT}\right)$, the r largest eigenvalues of $FN^{-1}\Lambda'\Lambda F'$. Then, we have $\left(B_{NT} + A_{NT}C_{NT}^{-1}\right)\Upsilon_{NT} = \Upsilon_{NT}\tilde{V}_{NT}$, where Υ_{NT} is the collection of unit length eigenvectors of the matrix $B_{NT} + A_{NT}C_{NT}^{-1}$. By the Assumption 1, we have $\left(N^{-1}\Lambda'\Lambda\right) \xrightarrow{p} \Sigma_\Lambda$ and $\left(D_{1T}^{-1}F'F\right) \xrightarrow{d} \Sigma_F$. Also, we have $A_{NT} = o_P(1)$. Furthermore, we may show that $C_{NT}^{-1} = O_P(1)$.

Since $F'\tilde{F}D_{1T}^{-2}$ is asymptotically block diagonal, we have C_{NT} as an asymptotically diagonal matrix. We may show that $\lim_{T,N\to\infty} C_{NT,ii} \neq 0$, in probability, for $i = 1, \ldots r$, where $C_{NT,ii}$ is the ii^{th} term of the matrix C_{NT}. Consider $(C_{NT})'(C_{NT})$.

$$\begin{aligned} C_{NT}'C_{NT} &= D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}F'\tilde{F}D_{1T}^{-2} = D_{1T}^{-2}\tilde{F}'\left(\frac{XX'}{N}\right)\tilde{F}D_{1T}^{-2} + o_P(1) \\ &= \tilde{V}_{NT} + o_P(1), \end{aligned}$$

where the second and third equalities followed by the Lemma (B.2) (i) and (ii). Using Lemma (B.2), we have $\tilde{V}_{NT} = O_P(1)$ and $\tilde{V}_{NT}^{-1} = O_P(1)$, which implies that $\lim_{N,T\to\infty}\tilde{V}_{NT,ii} \neq 0$, in probability. Then, $\lim_{N,T\to\infty} C_{NT,ii}^2 \neq 0$, in probability for $i = 1, \ldots r$. Thus, we have C_{NT} is bounded away from zero, $C_{NT}^{-1} = O_P(1)$. Hence, together with $A_{NT} = o_P(1)$, we obtain $B_{NT} + A_{NT}C_{NT}^{-1}$ converges, in distribution, to $B = \Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$.

Furthermore, by Assumption $1(v)$, we assume that the eigenvalues of $\Sigma_\Lambda\Sigma_F$ are distinct. Then, the eigenvalues of $B_{NT} + A_{NT}C_{NT}^{-1}$ are also distinct for large N and T. This implies that the eigenvectors of the $B_{NT} + A_{NT}C_{NT}^{-1}$ are unique except for the fact that these eigenvectors can be replaced by their negative (other sign) of themselves. Since C_{NT} and $\Upsilon_{NT} = C_{NT}\tilde{V}_{NT}^{\dagger -1/2}$ are functions of $\tilde{F}$, given the column sign of $\tilde{F}$, Υ_{NT} is uniquely determined. Using the eigenvalue perturbation theory, we have a unique eigenvector matrix Υ of B such that $\Upsilon_{NT} \xrightarrow{d} \Upsilon$. Since $\left(N^{-1}\Lambda'\Lambda\right)$ is positive definite, we can rewrite equation (B.17) as $D_{1T}^{-2}\tilde{F}'F = C_{NT}'\left(N^{-1}\Lambda'\Lambda\right)^{-1/2} = \tilde{V}_{NT}^{\dagger 1/2}\Upsilon_{NT}'\left(N^{-1}\Lambda'\Lambda\right)^{-1/2}$. Thus, together with $\tilde{V}_{NT}^{\dagger} \xrightarrow{d} V$ in Lemma (B.2), we have

$$D_{1T}^{-2}\tilde{F}'F \xrightarrow{d} V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2} := Q.$$

■

Lemma 3. *Suppose that Assumptions 1–4 hold. Then, as $N, T \to \infty$ with $\sqrt{N}||D_{1T}^{-2}|| \to 0$ for each given t, we have $\sqrt{N}\left(\tilde{F}_t - HF_t\right) \xrightarrow{d} V^{-1}QN\left(0, \Gamma_t\right) \stackrel{d}{=} N(0, \Sigma_{\tilde{F}})$, where Q is defined in Lemma (2), Γ_t is defined in Assumption 3, and Q is independent of $N(0, \Gamma_t)$.*

Proof of Lemma 3. Using the identity for $\tilde{F}, \tilde{\Lambda}$ and H as in the proof of Lemma (1), we obtain

$$\begin{aligned}\tilde{F}_t - HF_t &= \tilde{V}_{NT}^{-1}\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right)\\ &= \tilde{V}_{NT}^{-1}\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right).\end{aligned}$$

Using Lemma (B.2), we have $\tilde{V}_{NT}^{-1} = O_P(1)$. Furthermore, in Lemma (B.3), we have shown that $A_{1t} = O_P\left(\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right), A_{2t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right), A_{3t} = O_P\left(N^{-1/2}\right)$ and $A_{4t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right)$. Then, we have,

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = O_P\left(\sqrt{N}\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right) + O_P\left(\delta_{NT}^{-1}\right) + O_P(1).$$

Since $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$, we consider the following two cases:

Case 1. *If $O\left(N^{-1/2}\right) > O(\left\|D_{1T}^{-1}\right\|)$, we have $O\left(\delta_{NT}^{-1}\right) = O\left(N^{-1/2}\right)$, then,*

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = O_P(\left\|D_{1T}^{-1}\right\|) + O_P\left(N^{-1/2}\right) + O_P(1).$$

Case 2. *If $O(\left\|D_{1T}^{-1}\right\|) > O\left(N^{-1/2}\right)$, we have $O\left(\delta_{NT}^{-1}\right) = O\left(\left\|D_{1T}^{-1}\right\|\right)$, then,*

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = O_P(\sqrt{N}\left\|D_{1T}^{-2}\right\|) + O_P\left(\left\|D_{1T}^{-1}\right\|\right) + O_P(1).$$

Thus, for both cases, limiting distribution of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$ is determined by the third term under the condition $\lim_{N,T\to\infty}(N^{1/2}\left\|D_{1T}^{-2}\right\|) \to 0$.

Therefore, we have

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = \tilde{V}_{NT}^{-1}N^{1/2}D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\frac{F_s'\Lambda' e_t}{N} + o_P(1) = \tilde{V}_{NT}^{-1}D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_sF_s'\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_P(1).$$

By Assumption 3(ii), $N^{-1/2}\sum_{i=1}^{N}\lambda_i e_{it} \xrightarrow{d} N\left(0, \Gamma_t\right)$. Hence, together with Lemma (2) and Lemma (B.2), we have, $\sqrt{N}\left(\tilde{F}_t - HF_t\right) \xrightarrow{d} V^{-1}QN(0, \Gamma_t)$, where $\Gamma_t = \lim_{N,T\to\infty} N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\lambda_i\lambda_j' e_{it}e_{jt}\right)$. Since Q, the limiting distribution of $D_{1T}^{-2}\tilde{F}'F$, is determined only by the common factors, Q is independent of $N(0, \Gamma_t)$. ■

Lemma B. 4. *Suppose that Assumptions 1-6 are satisfied. Then, $D_{1T}^{-2}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\epsilon_{t+h} = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$ where $\delta_{NT}^{-1} = max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$.*

Proof. First, recall the identity (B.10),

$$\tilde{F}_t - HF_t = \tilde{V}_{NT}^{-1}\left\{D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\}.$$

Using the above identity, we have

$$D_{1T}^{-2}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\epsilon_{t+h} = \tilde{V}_{NT}^{-1}\left(I + II + III + IV\right), \tag{B.18}$$

where $I = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\epsilon_{t+h}$, $II = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\epsilon_{t+h}$, $III = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\epsilon_{t+h}$ and $IV = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\epsilon_{t+h}$.

According to Lemma (B.2), we have $\tilde{V}_{NT}^{-1} = O_P(1)$, and we may show that

$$I = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right), \quad II = O_P\left(N^{-1/2}T^{-1/2}\right), \quad III = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right), \quad IV = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right).$$

Consider each term of equation (B.18) separately.

$$I = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\epsilon_{t+h} = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}\epsilon_{t+h} + D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}HF_s\gamma_{st}\epsilon_{t+h} = I_1 + HI_2 \quad \text{(say)}.$$

Using Cauchy Schwarz inequality,

$$\begin{aligned}
||I_1|| &= \left\|D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}\epsilon_{t+h}\right\| = \left\|D_{1T}^{-4}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\left(\sum_{t=1}^{T-h}\gamma_{st}\epsilon_{t+h}\right)\right\| \\
&\leqslant \left\|D_{1T}^{-4}\right\|\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\left|\sum_{t=1}^{T-h}\gamma_{st}\epsilon_{t+h}\right|^2\right)^{1/2} \\
&\leqslant T^2\left\|D_{1T}^{-4}\right\|\frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T-h}|\gamma_{st}|^2\frac{1}{T}\sum_{t=1}^{T-h}|\epsilon_{t+h}|^2\right)^{1/2} \\
&= T^2\left\|D_{1T}^{-4}\right\|\frac{1}{\sqrt{T}}O_P\left(\delta_{NT}^{-1}\right),
\end{aligned}$$

since $\left(T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right) = O_P(\delta_{NT}^{-2})$ from Lemma (1), $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T-h}|\gamma_{st}|^2 \leqslant M$ from Lemma (B.1), and $T^{-1}\sum_{t=1}^{T-h}E|\epsilon_{t+h}|^2 = O(1)$ by Assumption 6. Therefore, together with the fact that $T||D_{1T}^{-2}|| = O(1)$, we have $\|I_1\| = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$.

Now, consider the second part of I. Using Cauchy-Schwarz inequality,

$$\begin{aligned}
E\|I_2\| = E\left\|D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}F_s\gamma_{st}\epsilon_{t+h}\right\| &\leqslant \left\|D_{1T}^{-4}\right\|\sum_{t=1}^{T-h}\sum_{s=1}^{T}E\left\|F_s\gamma_{st}\epsilon_{t+h}\right\| \\
&\leqslant \left\|D_{1T}^{-4}\right\|\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\gamma_{st}|\left(E\|F_s\|^2\right)^{1/2}\left(E|\epsilon_{t+h}|^2\right)^{1/2} = O\left(\left\|D_{1T}^{-2}\right\|\right),
\end{aligned}$$

as we have $E\left\|F_s\right\|^2 \leqslant M$ and $E\left|\epsilon_{t+h}\right|^2 \leqslant M$, and $T^{-1}\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\gamma_{st}| \leqslant M$ for some finite constant M by the Assumptions. Hence, $I = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(\left\|D_{1T}^{-2}\right\|\right) = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$ as $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$.

Now, we may show that $II = O_P\left(N^{-1/2}T^{-1/2}\right)$.

$$II = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\epsilon_{t+h} = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\epsilon_{t+h} + D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}HF_s\zeta_{st}\epsilon_{t+h} = II_1 + HII_2.$$

Using Cauchy Schwarz inequality and Assumption 5(i), we have

$$\begin{aligned}
||II_1|| &= \left\|D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\epsilon_{t+h}\right\| \leqslant \left\|D_{1T}^{-4}\right\|\left\|\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\sum_{t=1}^{T-h}\zeta_{st}\epsilon_{t+h}\right\| \\
&\leqslant \left\|D_{1T}^{-4}\right\|\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\left\|\sum_{t=1}^{T-h}\zeta_{st}\epsilon_{t+h}\right\|^2\right)^{1/2} \\
&= T^2\left\|D_{1T}^{-4}\right\|\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\underbrace{\frac{1}{T}\sum_{s=1}^{T}\left|\frac{1}{T}\sum_{t=1}^{T-h}\zeta_{st}\epsilon_{t+h}\right|^2}_{II_3(say)}\right)^{1/2} \\
&= T^2\left\|D_{1T}^{-4}\right\|\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}(II_3)^{1/2},
\end{aligned}$$

where we have

$$E(II_3) \leqslant \frac{1}{T}\sum_{s=1}^{T}E\left|\frac{1}{T}\sum_{t=1}^{T-h}\zeta_{st}\epsilon_{t+h}\right|^2 = \frac{1}{T}\sum_{s=1}^{T}E\left|\frac{1}{T}\sum_{t=1}^{T-h}\frac{1}{N}\sum_{i=1}^{N}\left(e_{is}e_{it} - E(e_{is}e_{it})\right)\epsilon_{t+h}\right|^2 = O\left(\frac{1}{NT}\right).$$

Then, using Lemma (1), we obtain, $||II_1|| = O(1)O_P(\delta_{NT}^{-1})O_P\left(N^{-1/2}T^{-1/2}\right) = O_P\left(N^{-1/2}T^{-1/2}\delta_{NT}^{-1}\right)$.

By Cauchy Schwarz inequality, Assumption 1, and the fact that $E(II_3) = O\left(N^{-1}T^{-1}\right)$, we have,

$$\begin{aligned}
\|II_2\| &= \left\|D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}F_s\zeta_{st}\epsilon_{t+h}\right\| \leqslant \left\|D_{1T}^{-3}\right\|\left\|\sum_{t=1}^{T-h}\sum_{s=1}^{T}D_{1T}^{-1}F_s\zeta_{st}\epsilon_{t+h}\right\| \\
&= \left\|D_{1T}^{-3}\right\|\left\|\sum_{s=1}^{T}D_{1T}^{-1}F_s\sum_{t=1}^{T-h}\zeta_{st}\epsilon_{t+h}\right\| \leqslant \left\|D_{1T}^{-3}\right\|\left(\sum_{s=1}^{T}\left\|D_{1T}^{-1}F_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\left\|\sum_{t=1}^{T-h}\zeta_{st}\epsilon_{t+h}\right\|^2\right)^{1/2} \\
&= O\left(\left\|D_{1T}^{-3}\right\|\right)O_P(1)O_P\left(\frac{T^{3/2}}{\sqrt{NT}}\right) = O_P\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}$$

Thus, together with Lemma (B.2) and $H = O_P(1)$, we obtain

$$II = II_1 + HII_2 = O_P\left(N^{-1/2}T^{-1/2}\delta_{NT}^{-1}\right) + O_P(1)O_P\left(N^{-1/2}T^{-1/2}\right) = O_P\left(N^{-1/2}T^{-1/2}\right).$$

Similar to the first and second terms, we can decompose the third term in equation (B.18) as follows:

$$III = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\epsilon_{t+h} = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\epsilon_{t+h} + D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}HF_s\eta_{st}\epsilon_{t+h} = III_1 + HIII_2.$$

Using Cauchy Schwarz inequality,

$$\begin{aligned}\|III_1\| &= \left\|D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\epsilon_{t+h}\right\| \leqslant \|D_{1T}^{-4}\|\left\|\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\sum_{t=1}^{T-h}\eta_{st}\epsilon_{t+h}\right\| \\ &\leqslant \|D_{1T}^{-4}\|\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\underbrace{\sum_{s=1}^{T}\left\|\sum_{t=1}^{T-h}\eta_{st}\epsilon_{t+h}\right\|^2}_{III_3(say)}\right)^{1/2},\end{aligned}$$

where we have

$$\begin{aligned}E(III_3) &\leqslant \sum_{s=1}^{T}\left(E\left\|\sum_{t=1}^{T-h}\eta_{st}\epsilon_{t+h}\right\|^2\right) = \sum_{s=1}^{T}E\left\|\sum_{t=1}^{T-h}\frac{F_s'\Lambda' e_t}{N}\epsilon_{t+h}\right\|^2 \\ &= \sum_{s=1}^{T}E\left\|F_s'\sum_{t=1}^{T-h}\frac{\Lambda' e_t\epsilon_{t+h}}{N}\right\|^2 \leqslant \sum_{s=1}^{T}E\|F_s\|^2\left(E\left\|\sum_{t=1}^{T-h}\frac{\Lambda' e_t\epsilon_{t+h}}{N}\right\|^2\right) = O\left(\frac{T^2}{N}\right),\end{aligned}$$

since $E\|F_s\|^2 = O(1)$ and $E\left\|N^{-1/2}T^{-1/2}\sum_{t=1}^{T-h}\sum_{i=1}^{N}\lambda_i e_{it}\epsilon_{t+h}\right\|^2 = O(1)$ by Assumptions 1 and 5.

Therefore, using Lemma (1) and the fact that $T\left\|D_{1T}^{-2}\right\| = O(1)$, we have,

$$\|III_1\| = O_P\left(||D_{1T}^{-4}||\sqrt{T}\delta_{NT}^{-1}TN^{-1/2}\right) = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right).$$

Again, using Cauchy Schwartz inequality, $T\left\|D_{1T}^{-2}\right\| = O(1)$, and Assumptions 1 and 5, we have,

$$\begin{aligned}III_2 &= \left\|D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}F_s\eta_{st}\epsilon_{t+h}\right\| = \left\|D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}F_sF_s'\frac{\Lambda' e_t}{N}\epsilon_{t+h}\right\| \\ &\leqslant \|D_{1T}^{-2}\|\left\|\sum_{t=1}^{T-h}\sum_{s=1}^{T}D_{1T}^{-2}F_sF_s'\frac{\Lambda' e_t\epsilon_{t+h}}{N}\right\| = \|D_{1T}^{-2}\|\left\|\left(\sum_{s=1}^{T}D_{1T}^{-2}F_sF_s'\right)\left(\sum_{t=1}^{T-h}\frac{\Lambda' e_t\epsilon_{t+h}}{N}\right)\right\| \\ &\leqslant \|D_{1T}^{-2}\|\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_sF_s'\right\|\left\|\sum_{t=1}^{T-h}\frac{\Lambda' e_t\epsilon_{t+h}}{N}\right\| = \left\|D_{1T}^{-2}\right\|O_P(1)O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right).\end{aligned}$$

Therefore, we have, $III = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right) = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right).$

Again from the decomposition, we have

$$IV = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\epsilon_{t+h} = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\xi_{st}\epsilon_{t+h} + D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}HF_s\xi_{st}\epsilon_{t+h} = IV_1 + HIV_2.$$

Consider each term separately. Using Cauchy-Schwartz inequality,

$$\|IV_1\| = \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \xi_{st} \epsilon_{t+h} \right\| \leqslant \left\| D_{1T}^{-4} \right\| \left\| \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \sum_{t=1}^{T-h} \xi_{st} \epsilon_{t+h} \right\|$$

$$\leqslant \left\| D_{1T}^{-4} \right\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\underbrace{\sum_{s=1}^{T} \left\| \sum_{t=1}^{T-h} \xi_{st} \epsilon_{t+h} \right\|^2}_{IV_3 \text{ (say)}} \right)^{1/2}.$$

From Lemma (B.1), and Assumptions 1 and 6, we obtain

$$E(IV_3) = E\left(\sum_{s=1}^{T} \left\| \sum_{t=1}^{T-h} \frac{F_t' \Lambda' e_s}{N} \epsilon_{t+h} \right\|^2 \right) = E\left(\sum_{s=1}^{T} \left\| \left(\frac{\Lambda' e_s}{N} \right)' \sum_{t=1}^{T-h} F_t \epsilon_{t+h} \right\|^2 \right)$$

$$\leqslant \sum_{s=1}^{T} E \left\| \frac{\Lambda' e_s}{N} \right\|^2 E \left\| \sum_{t=1}^{T-h} F_t \epsilon_{t+h} \right\|^2 \leqslant \sum_{s=1}^{T} E \left\| \frac{\Lambda' e_s}{N} \right\|^2 \sum_{t=1}^{T-h} E \left\| F_t \epsilon_{t+h} \right\|^2$$

$$\leqslant \sum_{s=1}^{T} E \left\| \frac{\Lambda' e_s}{N} \right\|^2 \sum_{t=1}^{T-h} E \left\| F_t \right\|^2 E \left\| \epsilon_{t+h} \right\|^2 = O\left(\frac{T^2}{N} \right).$$

Therefore, together with the fact that $T \left\| D_{1T}^{-2} \right\| = O(1)$ and Lemma (1), we have,

$$\|IV_1\| = O\left(\left\| D_{1T}^{-4} \right\| \right) O_P\left(T^{1/2} \delta_{NT}^{-1} \right) O_P\left(T N^{-1/2} \right) = O_P\left(N^{-1/2} \left\| D_{1T}^{-1} \right\| \delta_{NT}^{-1} \right).$$

Again, using Cauchy-Schwartz inequality and Assumptions 1, 3 and 6, we have

$$\|IV_2\| = \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_s \xi_{st} \epsilon_{t+h} \right\| = \left\| D_{1T}^{-2} \left(D_{1T}^{-1} \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right) \left(\sum_{t=1}^{T-h} D_{1T}^{-1} F_t \epsilon_{t+h} \right) \right\|$$

$$\leqslant \left\| D_{1T}^{-2} \right\| \left\| D_{1T}^{-1} \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right\| \left\| \sum_{t=1}^{T-h} D_{1T}^{-1} F_t \epsilon_{t+h} \right\|$$

$$\leqslant \left\| D_{1T}^{-2} \right\| \left\| D_{1T}^{-1} \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right\| \left(\sum_{t=1}^{T-h} \left\| D_{1T}^{-1} F_t \right\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \left| \epsilon_{t+h} \right|^2 \right)^{1/2}$$

$$= O_P\left(\left\| D_{1T}^{-2} \right\| \frac{\sqrt{T}}{\sqrt{N}} \right) = O_P\left(\frac{\left\| D_{1T}^{-1} \right\|}{\sqrt{N}} \right).$$

Thus, we have $IV = O_P\left(N^{-1/2} \left\| D_{1T}^{-1} \right\| \delta_{NT}^{-1} \right) + O_P\left(\left\| D_{1T}^{-1} \right\| N^{-1/2} \right) = O_P\left(\left\| D_{1T}^{-1} \right\| N^{-1/2} \right)$.
Since $\tilde{V}_{NT}^{-1} = O_P(1)$, Lemma (B.2), we have,

$$D_{1T}^{-2} \sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t \right) \epsilon_{t+h} = \tilde{V}_{NT}^{-1} \left\{ O_P\left(T^{-1/2} \delta_{NT}^{-1} \right) + O_P\left(N^{-1/2} T^{-1/2} \right) + O_P\left(N^{-1/2} \left\| D_{1T}^{-1} \right\| \right) \right\}$$

$$= O_P\left(T^{-1/2} \delta_{NT}^{-1} \right).$$

Therefore, we have shown $D_{1T}^{-1} \sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t \right) \epsilon_{t+h} \xrightarrow{p} 0$ as $T, N \to \infty$. ■

Lemma B. 5. *Let Assumptions 1-6 hold. If, in addition, $T/N \to 0$, then as $T, N \to \infty$, we have* $D_{1T}^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(\tilde{F}_t - HF_t\right)'(H^{-1})'\theta \xrightarrow{p} 0$.

Proof. By replacing $\hat{L}_t = \begin{pmatrix} \tilde{F}_t' & W_t' \end{pmatrix}'$, we obtain,

$$D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta = D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t W_t \end{pmatrix}\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta$$

$$= D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t - HF_t + HF_t \\ W_t \end{pmatrix}\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta$$

$$= \begin{pmatrix} D_{1T}^{-1}\sum_{t=1}^{T-h}(\tilde{F}_t - HF_t + HF_t)\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\ \\ \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \end{pmatrix}$$

$$= \begin{pmatrix} D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\ \\ 0 \end{pmatrix}$$

$$+ \begin{pmatrix} D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\ \\ \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \end{pmatrix} = \begin{pmatrix} A_1 + A_2 \\ A_3 \end{pmatrix}\left(H^{-1}\right)'\theta \qquad \text{(say)},$$

where $A_1 = D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)'$, $A_2 = D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)'$, and $A_3 = T^{-1}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'$.

By considering each term separately, we shall show that A_1, A_2, and A_3 converge to 0, in probability, as $N, T \to \infty$ with $T/N \to 0$.

First, recall the equation (B.10).

$$\tilde{F}_t - HF_t = \tilde{V}_{NT}^{-1}\left\{D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\},$$
$$= \tilde{V}_{NT}^{-1}\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right),$$

where $A_{1t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}$, $A_{2t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}$, $A_{3t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}$ and $A_{4t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}$.

Then, we may write A_1 as follows:

$$\begin{aligned} A_1 &= D_{1T}^{-1} \sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t\right) \left(\tilde{F}_t - HF_t\right)' \\ &= D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{t=1}^{T-h} \left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right) \left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)' \tilde{V}_{NT}^{-1} \\ &= D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} + D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1}. \end{aligned}$$

Using the triangle inequality,

$$\begin{aligned} \|A_1\| &= \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} + D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1} \right\| \\ &\leqslant \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} \right\| + \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1} \right\| \\ &= a_1 + a_2 \quad (say). \end{aligned}$$

Consider the two terms separately. Using Cauchy Schwartz inequality and Lemma (B.2), $\left\| \tilde{V}_{NT}^{-1} \right\| = O_P(1)$, we have

$$\begin{aligned} a_1 &= \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} \right\| \leqslant T \left\| D_{1T}^{-1} \right\| \left\| \tilde{V}_{NT}^{-1} \right\| \left\| \sum_{i=1}^{4} \frac{1}{T} \sum_{t=1}^{T-h} A_{it} A_{it}' \right\| \left\| \tilde{V}_{NT}^{-1} \right\| \\ &\leqslant T \left\| D_{1T}^{-1} \right\| \sum_{i=1}^{4} \frac{1}{T} \sum_{t=1}^{T-h} \|A_{it}\|^2 . \end{aligned}$$

In the proof of Lemma (1), we have shown that

$$\begin{aligned} T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{1t}\|^2 &= O_P\left(\left\|D_{1T}^{-2}\right\|\right), \quad & T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{2t}\|^2 &= O_P\left(N^{-1}\right), \\ T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{3t}\|^2 &= O_P\left(N^{-1}\right), \quad & T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{4t}\|^2 &= O_P\left(N^{-1}\right). \end{aligned}$$

Therefore, $a_1 = O\left(T \left\|D_{1T}^{-1}\right\|\right) \left\{O_P\left(\left\|D_{1T}^{-2}\right\|\right) + O_P\left(N^{-1}\right)\right\} = O_P\left(T \left\|D_{1T}^{-1}\right\| \delta_{NT}^{-2}\right)$.
Then, consider the cross terms such as $A_{it}A_{jt}'; i, j = 1, \ldots, 4$ for $i \neq j$. We may prove that the following cross term, a_2, is of order $o_P(1)$ under the condition $T/N \to 0$.

$$\begin{aligned} a_2 &= \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1} \right\| \leqslant T \left\| D_{1T}^{-1} \right\| \left\| \tilde{V}_{NT}^{-1} \right\| \left\| \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \frac{1}{T} \sum_{t=1}^{T-h} A_{it} A_{jt}' \right\| \left\| \tilde{V}_{NT}^{-1} \right\| \\ &\leqslant T \left\| D_{1T}^{-1} \right\| \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \left\| \frac{1}{T} \sum_{t=1}^{T-h} A_{it} A_{jt}' \right\| . \end{aligned}$$

Using Cauchy Schwarz inequality and the proof of Lemma (1), we have

$$\left\| T^{-1} \sum_{t=1}^{T-h} A_{1t} A_{2t}' \right\| \leqslant T^{-1} \left(\sum_{t=1}^{T-h} \|A_{1t}\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \|A_{2t}\|^2 \right)^{1/2}$$

$$= O\left(T^{-1}\right) O_P\left(\sqrt{T} \left\|D_{1T}^{-1}\right\|\right) O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right),$$

$$\left\| T^{-1} \sum_{t=1}^{T-h} A_{1t} A_{3t}' \right\| \leqslant T^{-1} \left(\sum_{t=1}^{T-h} \|A_{1t}\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \|A_{3t}\|^2 \right)^{1/2}$$

$$= O\left(T^{-1}\right) O_P\left(\sqrt{T} \left\|D_{1T}^{-1}\right\|\right) O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right),$$

$$\left\| T^{-1} \sum_{t=1}^{T-h} A_{1t} A_{4t}' \right\| \leqslant T^{-1} \left(\sum_{t=1}^{T-h} \|A_{1t}\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \|A_{4t}\|^2 \right)^{1/2}$$

$$= O\left(T^{-1}\right) O_P\left(\sqrt{T} \left\|D_{1T}^{-1}\right\|\right) O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right),$$

$$\left\| T^{-1} \sum_{t=1}^{T-h} A_{2t} A_{3t}' \right\| \leqslant T^{-1} \left(\sum_{t=1}^{T-h} \|A_{2t}\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \|A_{3t}\|^2 \right)^{1/2}$$

$$= T^{-1} O_P\left(\sqrt{\frac{T}{N}}\right) O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{1}{N}\right).$$

Similarly, we can show that the other cross terms $T^{-1} \sum_{t=1}^{T-h} A_{2t} A_{4t}'$ and $T^{-1} \sum_{t=1}^{T-h} A_{3t} A_{4t}'$ are also $O_P\left(N^{-1}\right)$. Then, as $\sqrt{T}/N \to 0$, we have,

$$a_2 = T \left\|D_{1T}^{-1}\right\| \left\{ O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right) + O_P\left(\frac{1}{N}\right) \right\} = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{\sqrt{T}}{N}\right) = o_P(1).$$

Together with the fact that $T \left\|D_{1T}^{-2}\right\| = O(1)$ and $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$,

$$A_1 = a_1 + a_2 = O_P\left(T \left\|D_{1T}^{-1}\right\| \delta_{NT}^{-2}\right) + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{\sqrt{T}}{N}\right) = o_P(1).$$

Then, consider the second term $A_2 = D_{1T}^{-1} \sum_{t=1}^{T-h} HF_t \left(\tilde{F}_t - HF_t\right)'$. By Cauchy Schwarz inequality, Assumption 1, Lemma (A.2) and Lemma (1), we obtain

$$\|A_2\| = \left\| D_{1T}^{-1} \sum_{t=1}^{T-h} HF_t \left(\tilde{F}_t - HF_t\right)' \right\| \leq ||H|| \left(\sum_{t=1}^{T-h} \left\|D_{1T}^{-1} F_t\right\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \left\|\tilde{F}_t - HF_t\right\|^2 \right)^{1/2}$$

$$= O_P\left(T^{1/2} \delta_{NT}^{-1}\right).$$

Since $\delta_{NT}^{-1} = \max[N^{-1/2}, ||D_{1T}^{-1}||]$, and $T \left\| D_{1T}^{-2} \right\| = O(1)$, we have $\delta_{NT}^{-1} = O\left(T^{-1/2}\right)$ as $T/N \to 0$. This implies, $\|A_2\| = O_P(1)$.

Thus, we cannot use this method to show $A_2 \xrightarrow{p} 0$ as $N, T \to \infty$ with $T/N \to 0$. Therefore, using $\tilde{F}_t - HF_t = \tilde{V}_{NT}^{-1} \{A_{1t} + A_{2t} + A_{3t} + A_{4t}\}$, we may rewrite A_2 as follows:

$$D_{1T}^{-1} \sum_{t=1}^{T-h} HF_t \left(\tilde{F}_t - HF_t\right)' = HD_{1T}^{-1} \sum_{t=1}^{T-h} F_t \left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)' \tilde{V}_{NT}^{-1}, \tag{B.19}$$

$$= H\left(B_1 + B_2 + B_3 + B_4\right) \tilde{V}_{NT}^{-1}, \tag{B.20}$$

where $B_1 = D_{1T}^{-1} \sum_{t=1}^{T-h} F_t A_{1t}'$, $B_2 = D_{1T}^{-1} \sum_{t=1}^{T-h} F_t A_{2t}'$, $B_3 = D_{1T}^{-1} \sum_{t=1}^{T-h} F_t A_{3t}'$, and $B_4 = D_{1T}^{-1} \sum_{t=1}^{T-h} F_t A_{4t}'$. We may consider each term separately.

$$\begin{aligned} B_1 &= D_{1T}^{-1} \sum_{t=1}^{T-h} F_t A_{1t}' = D_{1T}^{-1} \sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \gamma_{st} \right)' \\ &= D_{1T}^{-1} \sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2} \sum_{s=1}^{T} \gamma_{st}' \left(\tilde{F}_s - HF_s\right)' + D_{1T}^{-2} \sum_{s=1}^{T} \gamma_{st}' F_s' H' \right) = B_{11} + B_{12} H' \qquad \text{(say)} . \end{aligned}$$

Using Cauchy Schwarz inequality, Lemma (B.1), and Lemma (1) we have,

$$\begin{aligned} \|B_{11}\| &= \left\| D_{1T}^{-1} \sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2} \sum_{s=1}^{T} \gamma_{st}' \left(\tilde{F}_s - HF_s\right)' \right) \right\| \\ &\leqslant \left(\sum_{t=1}^{T-h} \left\| D_{1T}^{-1} F_t \right\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \underbrace{\left\| D_{1T}^{-2} \sum_{s=1}^{T} \gamma_{st}' \left(\tilde{F}_s - HF_s\right)' \right\|^2}_{b_{11t}} \right)^{1/2}, \end{aligned}$$

where we have,

$$b_{11t} \leqslant \left\| D_{1T}^{-2} \right\|^2 \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right) \left(\sum_{s=1}^{T} |\gamma_{st}|^2 \right) = O_p\left(\left\| D_{1T}^{-2} \right\| \delta_{NT}^{-2} \right)$$

Hence, we obtain, $\|B_{11}\| = O_P\left(\delta_{NT}^{-1}\right)$.

Consider $B_{12} = D_{1T}^{-1} \left(D_{1T}^{-2} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_t \gamma_{st}' F_s' \right)$. By Cauchy Schwarz inequality and Assumption $1(i)$, we obtain,

$$\begin{aligned} E \left\| D_{1T}^{-2} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_t \gamma_{st}' F_s' \right\| &\leqslant \left\| D_{1T}^{-2} \right\| \sum_{t=1}^{T-h} \sum_{s=1}^{T} E \left\| F_t \gamma_{st}' F_s' \right\| \\ &\leqslant \left\| D_{1T}^{-2} \right\| \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(E \left\| F_t \right\|^2 \right)^{1/2} \left(E \left\| F_s \right\|^2 \right)^{1/2} \left(E |\gamma_{st}|^2 \right)^{1/2} = O(1), \end{aligned}$$

since $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}| \leqslant M$ by Assumption 2, and $E\|F_t\|^2 \leqslant M$ by Assumption 1. Therefore, $B_{12} = O_p\left(\left\|D_{1T}^{-1}\right\|\right)$. Thus, we have

$$B_1 = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{1t}' = O_P\left(\delta_{NT}^{-1}\right) + O_P\left(\left\|D_{1T}^{-1}\right\|\right) = O_P\left(\delta_{NT}^{-1}\right). \tag{B.21}$$

Then, consider the second term in equation (B.19),

$$\begin{aligned} B_2 &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{2t}' = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right)' \\ &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right)' + D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T} HF_s\zeta_{st}\right)' \\ &= B_{21} + B_{22}H'. \end{aligned}$$

Using Cauchy Schwartz inequality, we have,

$$E\|B_{22}\| = E\left\|D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T} F_s'\zeta_{st}\right)\right\| \leqslant \left\|D_{1T}^{-3}\right\|\sum_{t=1}^{T-h}\left(E\|F_t\|^2\right)^{1/2}\left(E\left\|\sum_{s=1}^{T} F_s'\zeta_{st}\right\|^2\right)^{1/2}.$$

Using Assumption 1 and 2, we obtain, $E\left\|\sum_{s=1}^{T} F_s'\zeta_{st}\right\|^2 = O\left(TN^{-1}\right)$. Therefore, together with the fact that $T\left\|D_{1T}^{-2}\right\| = O(1)$, we have, $E\|B_{22}\| = O\left(N^{-1/2}\right)$. Again, using the Cauchy Schwartz inequality and Lemma (1), we have,

$$\begin{aligned} \|B_{21}\| &= \left\|D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right)'\right\| \\ &\leqslant \left\|D_{1T}^{-2}\right\|\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right\|^2\right)^{1/2} \\ &\leqslant \left\|D_{1T}^{-2}\right\|\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\zeta_{st}|^2\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2} \\ &= O_P\left(\left\|D_{1T}^{-2}\right\|\right)O_P\left(\frac{T}{\sqrt{N}}\sqrt{T}\delta_{NT}^{-1}\right) = O_P\left(\frac{\sqrt{T}}{\delta_{NT}\sqrt{N}}\right), \end{aligned}$$

since $E\left|N^{-1/2}\sum_{i=1}^{N}\left(e_{is}e_{it} - E(e_{is}e_{it})\right)\right|^4 \leqslant M$.

Together with Lemma (B.2), $\|H\| = O_P(1)$, we have

$$B_2 = O_P\left(\frac{\sqrt{T}}{\delta_{NT}\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right). \tag{B.22}$$

Similar to B_1 and B_2, we can rewrite B_3 as follows:

$$\begin{aligned} B_3 &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{3t}' = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right)' \\ &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right)' + D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T} HF_s\eta_{st}\right)' = B_{31} + B_{32}H'. \end{aligned}$$

Again, using Cauchy Schwartz inequality, Assumptions 1 and 3(iii), and Lemma (1), we obtain,

$$\begin{aligned} \|B_{31}\| &= \left\| D_{1T}^{-1}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\left(D_{1T}^{-2}\sum_{t=1}^{T-h}F_t\eta_{st}\right)\right\| \\ &\leqslant \left\|D_{1T}^{-1}\right\|\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\left\|D_{1T}^{-2}\sum_{t=1}^{T-h}F_t\eta_{st}\right\|^2\right)^{1/2} \\ &= O_p\left(T^{1/2}\delta_{NT}^{-1}N^{-1/2}\right). \end{aligned}$$

Again, using a similar argument, we have,

$$\|B_{32}\| = \left\|D_{1T}^{-1}\sum_{t=1}^{T-h}F_t\sum_{s=1}^{T}D_{1T}^{-2}F_s'\eta_{st}\right\| \leqslant \left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\eta_{st}\right\|^2\right)^{1/2} = O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right)$$

Since $H = O_P(1)$, Lemma (B.2), we have,

$$B_3 = B_{31} + B_{32}H' = O_P\left(\frac{\sqrt{T}}{\sqrt{N}\delta_{NT}}\right) + O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right) = O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right). \tag{B.23}$$

Similarly, we can show that $B_4 = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\right)$.

$$\begin{aligned} B_4 &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{4t}' = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right)' \\ &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\xi_{st}\right)' + D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T} HF_s\xi_{st}\right)' = B_{41} + B_{42}H'. \end{aligned}$$

Using Cauchy Schwartz inequality, Assumptions 1 and 3, Lemma (B.1), and the Lemma (1), we obtain,

$$B_{41} = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\frac{\Lambda' e_s F_t}{N}\right)' = \left(\sum_{t=1}^{T-h} D_{1T}^{-1}F_tF_t'D_{1T}^{-1}\right)\underbrace{\left(D_{1T}^{-1}\sum_{s=1}^{T}N^{-1}\Lambda' e_s\left(\tilde{F}_s - HF_s\right)\right)}_{B_{43}},$$

where we have,

$$\|B_{43}\| = \left\|D_{1T}^{-1}\right\|\left(\frac{1}{N^2}\sum_{s=1}^{T}\left\|\Lambda' e_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2} = \left\|D_{1T}^{-1}\right\| O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right)\left(\sqrt{T}\delta_{NT}^{-1}\right).$$

Hence, $B_{41} = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right)$.

Given Assumptions 1 and 3(iii), we obtain,

$$B_{42} = D_{1T}^{-1}\sum_{t=1}^{T-h}F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s\frac{\Lambda' e_s F_t}{N}\right)' = \sum_{t=1}^{T-h}D_{1T}^{-1}F_tF_t'D_{1T}^{-1}\left(D_{1T}^{-1}\sum_{s=1}^{T}\frac{\Lambda' e_s F_s'}{N}\right) = O_P\left(\frac{1}{\sqrt{N}}\right).$$

Therefore, we have

$$B_4 = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\right). \tag{B.24}$$

Thus, together with $\|H\| = O_P(1)$ and $\left\|\tilde{V}_{NT}^{-1}\right\| = O_P(1)$, from Lemma (B.2), and equations (B.21)-(B.24) we have

$$\begin{aligned} D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)' &= H\left(B_1 + B_2 + B_3 + B_4\right)\tilde{V}_{NT}^{-1} \\ &= O_P\left(\frac{1}{\delta_{NT}}\right) + \left(O_P\left(\frac{\sqrt{T}}{\sqrt{N}\delta_{NT}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right)\right) + O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right). \end{aligned}$$

Since $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$ and $T\left\|D_{1T}^{-2}\right\| = O(1)$, as $T, N \to \infty$ with $T/N \to 0$, we have $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|] = \max[N^{-1/2}, T^{-1/2}] = T^{-1/2}$. Hence, as $T, N \to \infty$ with $T/N \to 0$, we obtain,

$$D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)' = O_P\left(T^{-1/2}\right) + O_P\left(N^{-1/2}\right) + O_P\left(T^{1/2}N^{-1/2}\right) = o_P(1).$$

Now, consider the third term $A_3 = T^{-1}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'$. Using Cauchy Schwartz inequality and Lemma (1), we have

$$\|A_3\| = \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'\right\| \leq \left(\frac{1}{T}\sum_{t=1}^{T-h}\|W_t\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|\tilde{F}_t - HF_t\right\|^2\right)^{1/2} = O_P\left(\frac{\sqrt{T}}{\delta_{NT}}\right).$$

Thus, we cannot use this method to prove that $A_3 \xrightarrow{p} 0$ as $N, T \to \infty$ with $T/N \to 0$. Therefore, we consider the following method. Rewrite A_3 using A_{1t}, A_{2t}, A_{3t} and A_{4t} defined in Lemma (B.3),

$$A_3 = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)' = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)'\tilde{V}_{NT}^{-1} = \left(d_1 + d_2 + d_3 + d_4\right)'\tilde{V}_{NT}^{-1}, \tag{B.25}$$

where $d_1 = T^{-1}\sum_{t=1}^{T-h}W_tA_{1t}'$, $d_2 = T^{-1}\sum_{t=1}^{T-h}W_tA_{2t}'$, $d_3 = T^{-1}\sum_{t=1}^{T-h}W_tA_{3t}'$, and $d_4 = T^{-1}\sum_{t=1}^{T-h}W_tA_{4t}'$. Replacing A_{1t} by its definitions, we have

$$\begin{aligned} d_1 &= \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'H'\right) \\ &= d_{11} + d_{12}H'. \end{aligned}$$

By Cauchy- Schwartz inequality and Lemma (1), we have,

$$||d_{11}|| \leqslant \left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\gamma_{st}'\right\|^2\right)^{1/2} \tag{B.26}$$
$$= O_P\left(\sqrt{T}\right)O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right) = O_P(\delta_{NT}^{-1}),$$

since we assume that the observable series W_t are $I(1)$, and we have shown that $\left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\gamma_{st}'\right\|^2 = O_P\left(\left\|D_{1T}^{-2}\right\|\delta_{NT}^{-2}\right)$. Using Cauchy Schwartz inequality and Assumption 1, we have,

$$\|d_{12}\| = \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right)\right\| \leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right\|^2\right)^{1/2} = O_P\left(\left\|D_{1T}^{-1}\right\|\right).$$

Note that using Cauchy Schwartz inequality and Lemma (B.1), we have,

$$E\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right\|^2\right) \leqslant \sum_{t=1}^{T-h}E\left(\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right\|^2\right) \leqslant \left\|D_{1T}^{-2}\right\|^2\sum_{t=1}^{T-h}\left(\sum_{s=1}^{T}E\left\|F_s\right\|^2\right)\sum_{s=1}^{T}|\gamma_{st}|^2$$
$$= O\left(\left\|D_{1T}^{-2}\right\|\right).$$

Thus,

$$d_1 = O_P(\delta_{NT}^{-1}) + O_P(||D_{1T}^{-1}||) = O_P(\delta_{NT}^{-1}). \tag{B.27}$$

Now, consider the second term d_2.

$$d_2 = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s'\zeta_{st}'H'\right)$$
$$= d_{21} + d_{22}H'.$$

From Cauchy Schwartz inequality, Assumption 2(v), and Lemma (1), we obtain,

$$||d_{21}||^2 \leqslant ||D_{1T}^{-2}||^2\left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|\sum_{s=1}^{T}\zeta_{st}'\left(\tilde{F}_s - HF_s\right)\right\|^2\right)$$
$$= ||D_{1T}^{-2}||^2\left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)\left(\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\zeta_{st}|^2\right)\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)$$
$$= O_P\left(T||D_{1T}^{-2}||^2\right)O_P\left(\frac{T}{N}\right)O_P\left(T\delta_{NT}^{-2}\right) = O_P\left(\frac{T}{N\delta_{NT}^2}\right).$$
$$||d_{22}|| = \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_tD_{1T}^{-2}\sum_{s=1}^{T}F_s'\zeta_{st}'\right\| \leqslant \left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\zeta_{st}\right\|^2\right)^{1/2} = O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right).$$

Then, we have

$$d_2 = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(T^{1/2}N^{-1/2}\right) = O_P\left(T^{1/2}N^{-1/2}\right). \tag{B.28}$$

Now, consider the third term in equation (B.25). Similar with d_2, we can rewrite d_3 as follows:

$$d_3 = \frac{1}{T}\sum_{t=1}^{T-h} W_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h} W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h} W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s'\eta_{st}'H'\right)$$
$$= d_{31} + d_{32}H'.$$

Using Cauchy Schwartz inequality, we have,

$$\begin{aligned}\|d_{31}\| &= \left\|\frac{1}{T}\sum_{t=1}^{T-h} W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right)'\right\| \\ &\leqslant \left\|D_{1T}^{-2}\right\|\left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\eta_{st}|^2\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2} \\ &= O\left(\left\|D_{1T}^{-2}\right\|\right)O_P\left(\sqrt{T}\right)O_P\left(\frac{\left\|D_{1T}^{-1}\right\|^{-1}}{\sqrt{N}}\right)O_P\left(\sqrt{T}\delta_{NT}^{-1}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|^{-1}}{\sqrt{N}\delta_{NT}}\right).\end{aligned}$$

Note that from Lemma (B.1) and Assumption 1, we have,

$$\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\|\eta_{st}\|^2 = \frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left\|\frac{F_s'\Lambda' e_t}{N}\right\|^2 \leqslant \frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{\Lambda' e_t}{N}\right\|^2\sum_{s=1}^{T}\|F_s\|^2 = \frac{\left\|D_{1T}^{-2}\right\|^{-1}}{N}.$$

Since we assume that $W_t \sim I(1)$, and $\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s\eta_{st}\right\|^2 = O_P\left(N^{-1}\right)$, we have,

$$\|d_{32}\| = \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s\eta_{st}\right)\right\| \leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=}^{T}F_s\eta_{st}\right\|^2\right)^{1/2} = O_P\left(\frac{1}{\sqrt{N}}\right)$$

Thus,

$$d_3 = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|^{-1}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\right) = O_P\left(N^{-1/2}\right) \tag{B.29}$$

as $T, N \to \infty$ with $T/N \to 0$.

Similarly, we may show that $d_4 = O_P\left(N^{-1/2}\right)$.

$$d_4 = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\xi_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s\xi_{st}\right)'H'$$
$$= d_{41} + d_{42}H'$$

Using Cauchy Schwartz inequality and Lemma (1), we have,

$$\|d_{41}\| \leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s-HF_s\right)'\xi_{st}\right\|^2\right)^{1/2}$$

$$\leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\left\|D_{1T}^{-1}\right\|^2\underbrace{\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left\|D_{1T}^{-1}\xi_{st}\right\|^2}_{d_{43}}\sum_{s=1}^{T}\left\|\tilde{F}_s-HF_s\right\|^2\right)^{1/2},$$

where we have,

$$d_{43}=\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left\|D_{1T}^{-1}\frac{F_t'\Lambda'e_s}{N}\right\|^2 \leqslant \sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\sum_{s=1}^{T}\left\|\frac{\Lambda'e_s}{N}\right\|^2 = O_P\left(\frac{T}{N}\right).$$

We used Lemma (B.1) to bound d_{43}.

Thus, we have

$$d_{41}=O_P\left(1\right)O_P\left(\frac{\left\|D_{1T}^{-1}\right\|\sqrt{T}}{\sqrt{N}}\right)O_P\left(\sqrt{T}\delta_{NT}^{-1}\right)=O_P\left(\frac{\sqrt{T}}{\sqrt{N}\delta_{NT}}\right).$$

Again, by Cauchy Schwartz inequality and Assumptions 1 and 3(iii), we obtain,

$$\|d_{42}\|^2=\left\|\frac{1}{T}\sum_{t=1}^{T-h}W_tD_{1T}^{-1}\left(\sum_{s=1}^{T}\frac{D_{1T}^{-1}F_t'\Lambda'e_sF_s}{N}\right)\right\|^2 \leqslant \left\|D_{1T}^{-1}\right\|^2\left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)\left(\sum_{t=1}^{T-h}\left\|\sum_{s=1}^{T}\frac{D_{1T}^{-1}F_t'\Lambda'e_sF_s'}{N}\right\|^2\right)$$

$$\leqslant \left\|D_{1T}^{-1}\right\|^2\left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\left\|\sum_{s=1}^{T}\frac{\Lambda'e_sF_s}{N}\right\|^2$$

$$=O_P\left(\left\|D_{1T}^{-1}\right\|^2\right)O_P(1)O_P\left(\frac{\left\|D_{1T}^{-1}\right\|^{-2}}{N}\right)=O_P\left(\frac{1}{N}\right).$$

Hence,

$$||d_4||=O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right)+O_P\left(N^{-1/2}\right)=O_P\left(N^{-1/2}\right) \tag{B.30}$$

as $T,N\to\infty$ with $T/N\to 0$.

Therefore, together with Lemma (B.2) and equations (B.27)-(B.30), as $T/N\to 0$ for $T,N\to\infty$, we have,

$$A_3=(d_1+d_2+d_3+d_4)\,\tilde{V}_{NT}^{-1}=O_P(\delta_{NT}^{-1})+O_P\left(T^{1/2}N^{-1/2}\right)+O_P\left(N^{-1/2}\right)=o_P(1).$$

Hence, we have shown that

$$D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(\tilde{F}_t-HF_t\right)'(H^{-1})'\theta=\begin{pmatrix}A_1+A_2\\A_3\end{pmatrix}(H^{-1})'\theta=\begin{pmatrix}o_P(1)+o_P(1)\\o_P(1)\end{pmatrix}O_P(1)\text{Plim}(\hat{\theta})$$

$$\xrightarrow{p} 0 \qquad \text{as } N,T\to\infty \text{ with } T/N\to 0.$$

■

8 Appendix C

With the necessary lemmas listed in Appendix B, we are ready to prove the main theorems.

Theorem 1. *Suppose that Assumptions 1– 6 hold and that $T/N \to 0$. Let δ and the OLS estimator $\hat{\delta}$ be as in equation (9). Then, as $(N,T) \to \infty$, we have $D_T(\hat{\delta}-\delta) \xrightarrow{d} N\big(0, \Sigma_\delta\big)$, where $\Sigma_\delta = (\Psi')^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}\Sigma_L^{-1}\Psi^{-1}$, Σ_L and $\Sigma_{\epsilon L}$ are defined in Assumptions 1-6, and $H \oplus I \xrightarrow{d} \Psi$.*

Proof of Theorem 1. Let $D_T = [D_{1T} \oplus D_{2T}]$ where D_{1T} defined as $D_{1T} = diag[TI_p \oplus \sqrt{T}I_q]_{r\times r}$ and $D_{2T} = diag[T, ..., T]_m$. Then, the OLS estimator $\hat{\delta}$ of the forecasting model

$$Y_{t+h} = \hat{L}_t{}'\delta + \theta' H^{-1}\left(HF_t - \tilde{F}_t\right) + \epsilon_{t+h}$$

can be written as

$$\begin{aligned}
D_T(\hat{\delta}-\delta) &= \left(D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1}\right)^{-1}\left(D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\epsilon_{t+h} + D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(HF_t - \tilde{F}\right)' H^{-1}\theta\right) \\
&= \left(\underbrace{D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1}}_{Z}\right)^{-1}\left\{D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} HF_t \\ W_t \end{pmatrix}\epsilon_{t+h} + D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t - HF \\ 0 \end{pmatrix}\epsilon_{t+h}\right. \\
&\qquad\qquad \left. + D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t(HF_t - \tilde{F})'(H^{-1})'\theta\right\} \\
&= (Z)^{-1}(A+B+C) \qquad \text{(say)}
\end{aligned}$$

Then, we may consider each term separately.

$$\begin{aligned}
A &= D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} HF_t \\ W_t \end{pmatrix}\epsilon_{t+h} = D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} H & 0 \\ 0 & I_m \end{pmatrix}\begin{pmatrix} F_t \\ W_t \end{pmatrix}\epsilon_{t+h} \\
&= D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} H_1 & 0 & 0 \\ 0 & H_2 & 0 \\ 0 & 0 & I_m \end{pmatrix}\begin{pmatrix} E_t \\ G_t \\ W_t \end{pmatrix}\epsilon_{t+h} = (H_1 \oplus H_2 \oplus I_m)\, D_T^{-1}\sum_{t=1}^{T-h}L_t\epsilon_{t+h} \\
&\xrightarrow{d} \Psi\Sigma_{\epsilon L}^{1/2}N(0, I),
\end{aligned}$$

since $D_T^{-1}\sum_{t=1}^{T}L_t\epsilon_{t+h} \xrightarrow{d} \Sigma_{\epsilon L}^{1/2} \times N(0, I)$, using Assumption 6(iii) and $(H_1 \oplus H_2 \oplus I_m) = \Psi_0 \to \Psi$ as $N, T \to \infty$.

By Lemma (B.4), we have $B = D_{1T}^{-1} \sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t \right) \epsilon_{t+h} \xrightarrow{p} 0$ as $N, T \to \infty$, and by Lemma (B.5), if $T/N \to 0$ as $T, N \to \infty$, we have $C = D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_T (HF_t - \tilde{F}_t)'(H^{-1})'\theta \xrightarrow{p} 0$. [1]
Then, we may consider $Z = D_T^{-1} \left(\sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' \right) D_T^{-1}$. By writing $\hat{L}_t = \Psi_0 L_t + \hat{L}_t - \Psi_0 L_t$, we obtain,

$$
\begin{aligned}
&D_T^{-1} \left(\sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' \right) D_T^{-1} = D_T^{-1} \left(\sum_{t=1}^{T-h} (\Psi_0 L_t + \hat{L}_t - \Psi_0 L_t)(\Psi_0 L_t + \hat{L}_t - \Psi_0 L_t)' \right) D_T^{-1} \\
&= D_T^{-1} \left(\sum_{t=1}^{T-h} \left(\Psi_0 L_t L_t' \Psi_0' + \Psi_0 L_t \left(\hat{L}_t - \Psi_0 L_t \right)' \right) \right) D_T^{-1} \\
&+ D_T^{-1} \left(\sum_{t=1}^{T-h} \left(\left(\hat{L}_t - \Psi_0 L_t \right) (\Psi_0 L_t)' + \left(\hat{L}_t - \Psi_0 L_t \right) \left(\hat{L}_t - \Psi_0 L_t \right)' \right) \right) D_T^{-1} \\
&= \Psi_0 D_T^{-1} \left(\sum_{t=1}^{T-h} L_t L_t' \right) D_T^{-1} \Psi_0' + D_T^{-1} \left(\sum_{t=1}^{T-h} \left(\hat{L}_t - \Psi_0 L_t \right) L_t' \right) D_T^{-1} \Psi_0' \\
&+ \Psi_0 D_T^{-1} \left(\sum_{t=1}^{T-h} L_t \left(\hat{L}_t - \Psi_0 L_t \right)' \right) D_T^{-1} + D_T^{-1} \left(\sum_{t=1}^{T-h} \left(\hat{L}_t - \Psi_0 L_t \right) \left(\hat{L}_t - \Psi_0 L_t \right)' \right) D_T^{-1} \\
&\equiv z_1 + z_2 + z_3 + z_4 \qquad \text{(say)}.
\end{aligned}
$$

Then, we may show that $z_2 + z_3 + z_4 = o_P(1)$. First, consider z_2.

$$
\begin{aligned}
z_2 &= D_T^{-1} \left(\sum_{t=1}^{T-h} \left(\hat{L}_t - \Psi_0 L_t \right) L_t' \right) D_T^{-1} \Psi_0' \\
&= \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1} \left(\sum_{t=1}^{T-h} \begin{pmatrix} \tilde{F}_t - HF_t \\ 0 \end{pmatrix} \begin{pmatrix} F_t' & W_t' \end{pmatrix} \right) \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1} \begin{pmatrix} H & 0 \\ 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T-h} (\tilde{F}_t - HF_t) F_t' & \sum_{t=1}^{T-h} (\tilde{F}_t - HF_t) W_t' \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1} \begin{pmatrix} H & 0 \\ 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1} \begin{pmatrix} \left(\sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t \right) F_t' H D_{1T}^{-1} \right) & \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t \right) W_t' \right) \\ 0 & 0 \end{pmatrix}
\end{aligned}
$$

In the proof of Lemma (B.5), we have shown that $A_2 = \left(D_{1T}^{-1} \sum_{t=1}^{T-h} HF_t \left(\tilde{F}_t - HF_t \right)' \right) \xrightarrow{p} 0$, and $A_3 = \left(T^{-1} \sum_{t=1}^{T-h} W_t \left(\tilde{F}_t - HF_t \right)' \right) \xrightarrow{p} 0$ for $T, N \to \infty$ with $T/N \to 0$. Thus, $z_2 = o_P(1)$.

[1] If $T/N \to a (> 0)$, then $C \nrightarrow 0$ in probability as $N, T \to \infty$. Then there is a bias term as discussed in Goncalves (2014).

Also, $z_3 = z_2' = o_P(1)$. Now, consider the last term of Z.

$$z_4 = D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\hat{L}_t - \Psi_0 L_t\right)\left(\hat{L}_t - \Psi_0 L_t\right)'\right)D_T^{-1} = D_T^{-1}\left[\sum_{t=1}^{T-h}\begin{pmatrix}\tilde{F}_t - HF_t \\ 0\end{pmatrix}\begin{pmatrix}(\tilde{F}_t - HF_t)' & 0\end{pmatrix}\right]D_T^{-1}$$

$$= D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)' D_{1T}^{-1}.$$

Method 1

By the proof of Lemma (B.5), as $N, T \to \infty$ with $T/N \to 0$, we have

$$z_4 = A_1 D_{1T}^{-1} = O_P\left(T\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-2}\right)O_P\left(\left\|D_{1T}^{-1}\right\|\right) = O_P\left(\delta_{NT}^{-2}\right) \xrightarrow{p} 0.$$

Method 2: Similar arguments to the proof of Lemma B.7 of Choi (2017)

Using Lemma (3), asymptotic distribution of estimated factors, for $N, T \to \infty$ with $\sqrt{N}||D_{1T}^{-2}|| \to 0$, we have $(\tilde{F}_t - HF_t) = O_P\left(N^{-1/2}\right)$. Hence, $\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)' = O_P\left(N^{-1}\right)$. As $T\left\|D_{1T}^{-2}\right\| = O(1)$, we have, $D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)' D_{1T}^{-1} = O_P\left(\left\|D_{1T}^{-2}\right\|TN^{-1}\right) = o_P(1)$. Therefore, we have $z_4 = o_P(1)$. However, to follow this method, we need to have the condition $\sqrt{N}\left\|D_{1T}^{-2}\right\| \to 0$ also.

Thus, we obtain

$$Z = z_1 + o_P(1) = \Psi_0 D_T^{-1}\left(\sum_{t=1}^{T-h} L_t L_t'\right)D_T^{-1}\Psi_0' + o_P(1) \xrightarrow{d} \Psi\Sigma_L\Psi',$$

by Assumption 6(ii) and $\Psi_0 \xrightarrow{p} \Psi$ where Σ_L is a random matrix defined as in Assumptions. Then, together with the previous results for A, B, and C, we obtain

$$\begin{aligned} D_T(\hat{\delta} - \delta) &\xrightarrow{d} \left(\Psi\Sigma_L\Psi'\right)^{-1}\left\{\Psi\Sigma_{\epsilon L}^{1/2}N(0,I)\right\} \\ &\xrightarrow{d} \left(\Psi'\right)^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}^{1/2}N(0,I). \end{aligned}$$

Hence, $D_T(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_\delta)$ where $\Sigma_\delta = (\Psi')^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}\Sigma_L^{-1}\Psi^{-1}$.

■

Theorem 2. *Let Assumptions 1– 6 hold. Furthermore, suppose that* $\sqrt{N}||D_{1T}^{-2}|| \to 0$ *and* $T/N \to 0$ *as* $N, T \to \infty$*, and that* $(\hat{\Sigma}_\delta, \hat{\Sigma}_{\tilde{F}})$ *is a given consistent estimator of* $(\Sigma_\delta, \Sigma_{\tilde{F}})$*. Then, we have*

$$\frac{\hat{Y}_{T+h|T} - Y_{T+h|T}}{\sqrt{\hat{B}_T}} \xrightarrow{d} N(0,1) \quad \text{as } N, T \to \infty$$

where $\hat{B}_T = \left[\hat{L}_T D_T^{-1}\hat{\Sigma}_\delta D_T^{-1}\hat{L}_T' + N^{-1}\hat{\theta}'\hat{\Sigma}_{\tilde{F}}\hat{\theta}\right]$ *is a consistent estimator of the asymptotic variance, denoted* B_T*, of the conditional forecasting error that appears in the numerator.*

Proof of Theorem 2. Since an estimator of $Y_{T+h|T}$ is $\hat{Y}_{T+h|T} = \hat{\delta}'\hat{L}_T$, and

$$Y_{T+h|T} = \delta'\hat{L}_T + \theta' H^{-1}\left(HF_T - \tilde{F}_T\right),$$

we have,

$$\begin{aligned}
\hat{Y}_{T+h|T} - Y_{T+h|T} &= \left(\hat{\delta} - \delta\right)' \hat{L}_T + \theta' H^{-1}\left(\tilde{F}_T - HF_T\right) \\
&= \left(\hat{\delta} - \delta\right)' D_T D_T^{-1}\hat{L}_T + \frac{1}{\sqrt{N}}\theta' H^{-1}\sqrt{N}\left(\tilde{F}_T - HF_T\right) \\
&= \left(D_T\left(\hat{\delta} - \delta\right)\right)' D_T^{-1}\hat{L}_T + \frac{1}{\sqrt{N}}\theta' H^{-1}\sqrt{N}\left(\tilde{F}_T - HF_T\right).
\end{aligned}$$

Using Theorem (1), the limiting distribution of the estimated parameters, along with $T/N \to 0$, we have $D_T\left(\hat{\delta} - \delta\right) \xrightarrow{d} N(0, \Sigma_\delta)$.

By Lemma (3), the limiting distribution of the estimated factors for $\sqrt{N}||D_{1T}^{-2}|| \to 0$, we have

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) \xrightarrow{d} N(0, \Sigma_{\tilde{F}_t}).$$

Moreover, $D_T\left(\hat{\delta} - \delta\right)$ and $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$ are asymptotically independent as the limit of the first term determined by the regression errors and the limit of the second term determined by the idiosyncratic errors. Hence, the limiting distribution of the forecast error conditional on $\{L_t\}_{t=1}^T$ is

$$\hat{Y}_{T+h|T} - Y_{T+h|T} \xrightarrow{d} N\left(0, B_T\right),$$

where $B_T = \hat{L}_T' D_T^{-1}\Sigma_\delta D_T^{-1}\hat{L}_T + N^{-1}\theta'\Sigma_{\tilde{F}_T}\theta$.

Furthermore, $\hat{B}_T = \hat{L}_T' D_T^{-1}\hat{\Sigma}_\delta D_T^{-1}\hat{L}_T + N^{-1}\hat{\theta}'\hat{\Sigma}_{\tilde{F}_T}\hat{\theta}$ is the consistent estimator of the asymptotic variance of the forecast error where $\hat{\Sigma}_\delta$ and $\hat{\Sigma}_{\tilde{F}_T}$ defined as in the main paper. Therefore,

$$\frac{\hat{Y}_{T+h|T} - Y_{T+h|T}}{\sqrt{\hat{B}_T}} \xrightarrow{d} N(0,1) \text{ as } N, T \to \infty.$$

■

9 Appendix D: Simulation Results

In this section we present the simulation results on the coverage rates of residual based t-percentile bootstrap prediction intervals when the error distribution is normal and when it is t with 5 degrees of freedom.

Results when the error distribution is normal

The design of the simulation for this part is the same as that in section 3 of the main paper. Thus, the two DGPs are:

DGP1 : $\epsilon_t \sim N(0, 1)$

DGP2 : $\epsilon_t \sim N(0, 3^{-1}F_{2t}^2)$.

Table 5: Coverage rates (%) of residual based 95% bootstrap (t-percentile) prediction intervals for one-step ahead forecasts when the error distribution is normal

		$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
	T\N	30	50	100	200	30	50	100	200	30	50	100	200
DGP1	30	82	84	87	90	82	82	88	89	83	82	86	91
	50	83	86	88	87	82	87	89	86	82	87	90	86
	100	78	82	88	86	80	82	89	87	81	82	89	87
	200	85	84	84	88	85	84	85	89	86	84	85	88
DGP2	30	88	90	91	92	90	88	91	90	90	91	91	92
	50	90	92	93	93	89	93	93	94	91	94	94	95
	100	89	88	91	92	88	89	90	90	90	91	94	93
	200	86	87	93	90	86	86	93	89	89	88	94	92

Results when the error distribution is t_5

For this part of the simulation study, we used DGP1 as for the previous table, except that the errors $\{\epsilon_t\}$ were generated from t_5 distribution, instead of the normal distribution. Top panel of Table 6 provides the coverage rates of the 95% asymptotic prediction interval obtained assuming that the errors are normal. Bottom panel of Table 6 provides the coverage rates for the residual based 95% bootstrap t-percentile prediction interval.

Table 6: Coverage rates (%) of 95% prediction intervals for one-step ahead forecasts when the error distribution is t_5.

	$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
T\N	30	50	100	200	30	50	100	200	30	50	100	200
Forecast interval assuming that the errors are normally distributed												
30	96	95	97	95	97	95	97	96	97	95	96	95
50	94	93	94	96	94	93	94	96	93	94	93	97
100	95	93	96	97	95	93	96	97	95	93	96	97
200	94	96	94	95	94	96	94	95	94	97	94	96
Residual based bootstrap prediction interval												
30	84	84	87	88	84	84	86	87	84	84	85	87
50	82	84	87	90	84	84	88	90	85	85	87	89
100	83	83	88	91	84	84	89	91	84	84	88	92
200	83	88	83	86	84	89	84	86	85	90	84	87