MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

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

Forecasting a Nonstationary Time Series with a Mixture of Stationary and Nonstationary Factors as Predictors

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

May 2020

Working Paper 19/20

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS AMBA ACCREDITED

Forecasting a Nonstationary Time Series with a Mixture of Stationary and Nonstationary Factors as Predictors

Sium Bodha Hannadige[1], Jiti Gao, Mervyn J Silvapulle, and Param Silvapulle.

Department of Econometrics and Business Statistics, Monash University, Australia [2]

Abstract

This paper develops a method for forecasting a nonstationary time series, such as GDP, using a set of high-dimensional panel data as predictors. To this end, we use what is known as a factor augmented regression [FAR] model that contains a small number of estimated factors as predictors; the factors are estimated using time series data on a large number of potential predictors. The validity of this method for forecasting has been established when all the variables are stationary and also when they are all nonstationary, but not when they consist of a mixture of stationary and nonstationary ones. This paper fills this gap. More specifically, we develop a method for constructing an asymptotically valid prediction interval using the FAR model when the predictors include a mixture of stationary and nonstationary factors; we refer to this as *mixture-FAR* model. This topic is important because typically time series data on a large number of economic variables is likely to contain a mixture of stationary and nonstationary variables. In a simulation study, we observed that the mixture-FAR performed better than its competitor 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 proposed in this paper performed better than its aforementioned competitors.

JEL Classification: C22, C33, C38, C53.

Keywords: Bootstrap; generated factors; panel data; prediction interval.

[1]Corresponding author.

[2]Email addresses for authors: sium.bodhahannadige@monash.edu, jiti.gao@monash.edu, mervyn.silvapulle@monash.edu, and Param Silvapulle.

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 *factor model* for panel data and a *regression model* for predicting a time series are used jointly. 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 and Watson 1998a, 1998b, 2002b). The large number of economic variables that are potential predictors 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 and Perron (2007), Smeekes and Wijler (2019)). The objective of this paper is to develop new methodology 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 and Ng [2002, 2006], Gonçalves and Perron (2014)), and also when they are all nonstationary (Choi 2017), but not when there is 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 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) and Bai and 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 and Ng (2007), Stock and Watson (2012), Cheng and 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.

Next, 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 methodologically 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 it is common to find a mixture of stationary and nonstationary generated factors. For example, Bai (2004) studied employment fluctuations across 60 industries in the U.S 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 contains information to accurately identify the monetary transmission mechanism in the US. Eickmeier (2005) used a large-scale ($N > 300$) dynamic factor model and found euro-area economies share 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 and 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 and 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 and Perron (2007)) for generating a mixture of stationary and nonstationary factors. 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 and 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 methodology under the assumption $T/N \to 0$. We show the consistency and normality of estimators of the parameters of the forecast 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 developed in this paper 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 proposed in this paper performed better than its aforementioned competitors. Again, the results of this empirical study corroborate the general observation of our simulation study, namely, the method developed in this paper improves over the current methods for forecasting based on the FAR methodology.

The rest of this paper is organized as follows. Section 2 introduces the model and the assumptions, and obtains the asymptotic distribution of the OLS estimators and the prediction intervals. Section 3 reports the results of a simulation study. An empirical example using the FRED-QD data is presented in Section 4. Section 5 concludes. Several lemmas in Appendix A indicate the main steps for the proofs of the lemmas and theorems stated in Section 2; the proofs of the lemmas in Appendix A will be available in Supplementary Materials to this paper. Some of the simulation results are reported in Appendix B.

Our main focus is on forecasting a nonstationary variable, although the theory presented here is valid for forecasting a stationary variable as well with a mixture of stationary and nonstationary factors as predictors. However, this topic requires further attention, particularly on the empirical aspects, and hence will not be considered further in this paper.

Finally, let us introduce the following notation: Throughout this paper $A > 0$ denotes that the matrix is positive definite, and $X \oplus Y = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$ for the matrices X and Y; $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and in distribution, respectively.

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 structure of the two models that are used for predicting Y_{T+h} take the following form:

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

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

where FAR stands for *factor augmented regression*, 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 "diffusion index forecasting model" by Stock and Watson (2002a).

A point of departure of our paper from the current literature is that we allow the r factors 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, $F = (F_1, \ldots, F_T)'$ be the $T \times r$ matrix of unobservable common factors, $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ be the matrix of factor loadings, and $e = \left[e_{it}\right]_{T\times N}$ be the matrix of error terms from the factor model. Then the factor model (1) can also be expressed as

$$X = F\Lambda' + e. \tag{3}$$

Since the stationary and nonstationary terms need to 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,

$$E_t = E_{t-1} + u_t, \tag{4}$$

where u_t is stationary. Substituting $F_t' = (E_t', G_t')'$, the factor model (1) and the FAR model (2) take the form

$$X_{it} = \lambda_i^{(1)'} E_t + \lambda_i^{(2)'} G_t + e_{it}, \qquad (i = 1, ..., N; t = 1, ..., T) \tag{5}$$

$$Y_{t+h} = \alpha' E_t + \beta' G_t + \omega' W_t + \epsilon_{t+h}, \quad (t = 1, ..., T) \tag{6}$$

where $\lambda_i = (\lambda_i^{(1)'}, \lambda_i^{(2)'})'$ and $\theta = (\alpha', \beta')'$. As expected, estimates of the coefficients α and β of the non-stationary 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:

$$D_{1T} = \text{diag}(TI_p,\ T^{1/2}I_q)_{r\times r}, \qquad D_{2T} = TI_m \qquad D_T = \text{diag}(D_{1T},\ D_{2T}). \tag{7}$$

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 the panel Information Criteria (IC) introduced by Bai (2004) and Bai and Ng (2002).

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, the matrix F of common factors is estimated by $\tilde{F}$. 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. In this paper, we do not explore the effects of misclassification resulting from such pre-tests.

Once the factors have been estimated, the 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 are classified as nonstationary and their corresponding eigenvalues are in the decreasing order, and the remaining q columns are 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 $\tilde{V}_{NT} = diag(\tilde{V}_{p,NT}, \tilde{V}_{q,NT})$; then $\tilde{V}_{NT}$ is diagonal. 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

$$H = N^{-1}\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda. \tag{8}$$

If all the variables are stationary then the foregoing H reduces the expression in Bai and Ng (2002), and if all the variables are nonstationary then it reduces the form in Bai (2004) and Choi (2017).

Let $\hat{L}_t = (\tilde{F}_t', W_t)'$, and $\delta = (\theta' H^{-1}, \omega')'$; just as H, δ is also a function of the data and unknown population parameters. Then, the FAR model (6) can be written as

$$\begin{aligned} Y_{t+h} &= \theta' F_t + \omega' W_t + \epsilon_{t+h} = \theta' H^{-1}(HF_t - \tilde{F}_t + \tilde{F}_t) + \omega' W_t + \epsilon_{t+h} \\ &= \theta' H^{-1}\tilde{F}_t + \omega' W_t + \theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h} \\ &= \delta'\hat{L}_t + \theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h}. \end{aligned} \tag{9}$$

Let $(\hat{\alpha}, \hat{\beta}, \hat{\omega})$ denote the the ordinary least squares [OLS] estimator of (α, β, ω) obtained by regressing Y_{t+h} on $\hat{L}_t$ $(t = 1, \ldots, T-h)$. Then

$$\hat{\delta} = (\hat{\alpha}', \hat{\beta}', \hat{\omega}')' = \left(\sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t{}'\right)^{-1} \sum_{t=1}^{T-h} \hat{L}_t Y_{t+h}. \tag{10}$$

Later we will show that $\theta' H^{-1}\left(HF_t - \tilde{F}_t\right)$ is asymptotically centered at zero in the limit, and hence $\theta' H^{-1}\left(HF_t - \tilde{F}_t\right) + \epsilon_{t+h}$ could be treated as an error term centered at zero. 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 separately. 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 (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. Factors are not cointegrated.

iv. The loadings λ_i are either deterministic and $\|\lambda_i\| \leqslant M$ satisfying $\Lambda'\Lambda/N \to \Sigma_\Lambda$, 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 and r is known.

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 can 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 factors. To state the next assumption, let us introduce the following notation:

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

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. By allowing weak correlation among the idiosyncratic errors, we have an approximate factor structure which is more general than a strict factor model in which e_{it} are independent.

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 $t=1,\ldots,T$, we have $N^{-1/2}\Lambda' e_t \xrightarrow{d} N\left(0,\Gamma_t\right)$ as $N\to\infty$.

iii. $\left\|N^{-1/2}D_{1T}^{-1}\sum_{t=1}^{T}\Lambda' e_t F_t'\right\| = O_p(1)$, uniformly in N and T.

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 and 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 $\min\{N, ||D_{1T}^{-2}||^{-1}\}$. The next lemma shows the consistency of generated factors.

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*

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

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).

Assumptions 1 to 3 are sufficient to establish the consistency of generated factors. Next, we introduce another assumption to derive the asymptotic distributions of the estimated factors.

Assumption 4. Weak dependence of idiosyncratic errors

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

ii. $\sum_{j=1}^{N}|\tau_{ij}| \leqslant M$, for each $i = 1, ..., N$,

where γ_{st} and τ_{ij} are defined in Assumption 2.

2.3.2 Asymptotic distribution of the generated factors

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 $var\left(\tilde{F}_t - HF_t\right)$ 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

In this subsection, we obtain the limiting distribution of the estimators of the parameters in the FAR model (2). First, replace the unobservable common factor F by its estimator $\tilde{F}$, which results in (see, Equation 9),

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

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

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. $D_T^{-1}\sum_{t=1}^{T}L_tL_t'D_T^{-1} \xrightarrow{d} \Sigma_L$ 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.

Assumption 5 imposes restrictions on the degree of dependence, over time, among the idiosyncratic errors, 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. Parts (ii) and (iii) of Assumption 6 are used for establishing the asymptotic distribution of the estimated parameters.

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 (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 $H \oplus I \xrightarrow{d} \Psi$.*

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 $T, T^{1/2}, T$ respectively. Consequently, the limiting distribution in this theorem reduces to the following known corresponding results: (a) in Bai and Ng (2002) for the FAR model with $I(0)$ variables only, and (b) in 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 $T/N \to 0$ to ensure 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 $T/N \to c$ with $0 < c < \infty$, then the limiting normal distribution would have a nonzero mean, and hence the estimator would not be consistent. In fact, Gonçalves and 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

$$\hat{\Sigma}_{\hat{\delta}} = \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1} \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{\epsilon}_{t+h}^2 \hat{L}_t \hat{L}_t' D_T^{-1} \right) \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1}. \quad (11)$$

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

$$\hat{\Sigma}_{\hat{\delta}} = \hat{\sigma}_\epsilon^2 \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1}, \quad (12)$$

where $\hat{\sigma}_\epsilon^2 = T^{-1} \sum_{t=1}^{T-h} \hat{\epsilon}_{t+h}^2$, the estimated 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 (9) and (10). 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 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*

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

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.*

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

$$\hat{Y}_{T+h|T} - Y_{T+h|T} = \left(\hat{\delta} - \delta \right)' \hat{L}_T + \theta' H^{-1} \left(\tilde{F}_t - H F_t \right). \quad (13)$$

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 suggested in (11) or(12) depending on the assumptions. A consistent estimator of $\Sigma_{\tilde{F}_T}$ is [3]

$$\hat{\Sigma}_{\tilde{F}_T} = \tilde{V}_{NT}^{-1} \hat{\Gamma}_T \tilde{V}_{NT}^{-1}, \tag{14}$$

where $\hat{\Gamma}_T$, an $r \times r$ matrix, is an estimator of the asymptotic variance 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 (14), we need a feasible estimator of Γ_T. As suggested by Bai and Ng (2006), depending on the assumptions, $\hat{\Gamma}_t$ may take one of the following forms:

$$\hat{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N} \hat{e}_{it}^2 \tilde{\lambda}_i \tilde{\lambda}_i' \tag{15}$$

$$\hat{\Gamma}_t = \hat{\sigma}_e^2 \frac{1}{N}\sum_{i=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_i' \tag{16}$$

$$\hat{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_j' \hat{e}_{it}\hat{e}_{jt}, \tag{17}$$

where $\hat{e}_{it} = X_{it} - \tilde{\lambda}_i' \tilde{F}_t$. For cross sectionally uncorrelated idiosyncratic errors, the forms of $\hat{\Gamma}_t$ in (15) and (16) 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 (16) would be suitable. The form in (17) 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, we may construct a $100(1-\alpha)\%$ confidence interval for the conditional mean $Y_{T+h|T}$ as

$$\left(\hat{Y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B}_T}, \hat{Y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B}_T}\right), \tag{18}$$

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

$$\hat{\epsilon}_{T+h} = \hat{Y}_{T+h|T} - Y_{T+h} = \hat{L}_T'\left(\hat{\delta} - \delta\right) + \theta' H^{-1}\left(\tilde{F}_t - HF_t\right) - \epsilon_{T+h}. \tag{19}$$

Therefore, the limiting distribution of forecast error also depends on the distribution of the regression error ϵ_{t+h}. Let us suppose that $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, for all t. 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 $\hat{\sigma}_\epsilon^2 = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$.

[3] It follows from Lemma 2 that $var(\tilde{F}_t - HF_t) = V^{-1}Q\Gamma_t Q' V^{-1}$. We obtain the estimator in (14) by using $Q = D_{1T}^{-2}\tilde{F}'F$ and $\tilde{Q} = D_{1T}^{-2}\tilde{F}'\tilde{F}$ which is I_r by normalization condition.

Then, a 95% asymptotic prediction interval for Y_{T+h} is

$$\left(\hat{Y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B_T} + \hat{\sigma}_\epsilon^2}, \hat{Y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B_T} + \hat{\sigma}_\epsilon^2}\right). \tag{20}$$

3 Finite sample properties

This section reports the results of a simulation study that we carried out to evaluate the finite sample properties of the methods developed in this paper. The results are reported in two parts. In the first, we report the simulation results for (a) bias of the estimators, (b) coverage rates for 95% confidence intervals of the parameters $\{\alpha, \beta, \omega\}$ in the FAR model, and (c) coverage rates of 95% prediction intervals. In the second part, we compare the out-of-sample forecast performance of the method developed in this paper with nonstationary-FAR and AR(4).

Design of the simulation study

The data generating process [DGP] for the FAR :

$$\begin{aligned} Y_{t+1} &= \alpha F_{1t} + \beta F_{2t} + \omega Y_t + \epsilon_{t+h} \qquad (t = 1, \ldots, T-1) & (21) \\ F_{1t} &= F_{1,t-1} + v_t, \qquad \begin{pmatrix} v_t \\ F_{2,t} \end{pmatrix} \sim MVN\begin{pmatrix} 0, & C \end{pmatrix}, \quad C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. & (22) \end{aligned}$$

For ρ in (22), 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 is generated by

$$X_{it} = \lambda_i^{(1)} F_{1t} + \lambda_i^{(2)} F_{2t} + e_{it}, \tag{23}$$

with the λ_i's drawn from $N(0,1)$ and the error terms $\{e_{it}\}$ as listed 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 $\epsilon_t \sim N\left(0, 3^{-1}F_{2t}^2\right)$.

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, the error $var(\epsilon_t)$ in the FAR model varies over time and the error $var(e_{it})$ is distributed uniformly over $[0.5, 1.5]$; therefore, the average variance is the same as 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 (12) and (11) for DGP1 and {DGP2, DGP3} respectively.

Results:

The tabulated values of the simulation results are provided in Appendix C. Table 6 provides two sets of results for the case when the factors are independent. The first set in the top panel is the bias of the OLS estimator. As the set of estimated vector parameters does not converge to its true value but to a rotation of it, $H^{-1}\theta$ (each diagonal of H converges to ± 1), we report the mean bias of the rotated OLS coefficients over 5000 simulations. Note that this result is infeasible in the real-world application as the rotation matrix H depends on the population parameters such as true latent factors. The bottom panel of Table 6, reports the coverage rates for a 95% confidence intervals. Table 7 and Table 8 provide, the corresponding results for the case when two factors are correlated with 0.5 and 0.9 correlation coefficients respectively. The results in Table 6, Table 7, and Table 8 show that the dependency between the two factors did not make a significant difference to parameter estimation in the sense that bias and coverage rates in all three tables are same of the order of magnitude. The size of the bias appears to decrease as the panel size increases.

The bias as a proportion of the magnitude of the parameter, namely bias$(\hat{\theta})/|\theta|$, where θ may be α, β, or ω, are less than 6%, and are less than 2% in most cases. Thus, the bias appears to be small. The coverage rates of the 95% confidence intervals differ for α, β, and ω. For ω, the coverage rates are high; they are close to 99%. By contrast, the coverage rates for α and β are below the nominal level, but increase with T toward the nominal level. As expected, the coverage rate for α and β remain steady as N increases.

Table 5 reports the coverage rates of 95% prediction intervals for Y_{T+1}. These are based on the assumption that the regression errors are normal. These results show that the coverage rate increases with the panel dimension. We also evaluated the prediction intervals using the t-percentile bootstrap prediction intervals; this method does not assume that the errors are normal. The bootstrap intervals provided better coverage rates; the validity of the bootstrap is yet to be established.

3.1 Performance of mixture-FAR model for forecasting

In this subsection, we evaluate the performance of the mixture-FAR approach relative to nonstationary-FAR. 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

$$R^2_{os} = 1 - \left(\sum_{i=T_1+1}^{T} (Y_i - \hat{Y}_i)^2 \right) \left(\sum_{i=T_1+1}^{T} (Y_i - \tilde{Y}_i)^2 \right)^{-1}, \quad (24)$$

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 out-of-sample forecast performance 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 $t = T_1 + 1, \ldots, T$, 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.

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 with the corresponding more special case, nonstationary-FAR. Table 1 provides the results for this comparison. It is evident that the mixture-FAR method performs 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 1, the MSE of prediction for the 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.

Table 1: Calculated R^2_{os} of mixture FAR model compared to the 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

R^2_{os} for mixture FAR model relative to nonstationary FAR are calculate from one-step ahead out-of-sample forecasts. Here, the forecasting variable Y is nonstationary ($I(1)$).

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

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 based on the t-percentile bootstrap; the theory for the asymptotic validity of the bootstrap is not provided. We compare and contrast the out-of-sample forecasting performance of the mixture-FAR with the corresponding non-stationary-FAR and the AR(4) models. To quantify the out-of-sample forecasting performance, we use R_{os} defined in (24).

4.1 Data description

The data were collected from FRED-MD and FRED-QD that are well-known monthly and quarterly databases. The latter contains 246 US macroeconomic time series for the period 1959:Q1 to 2018:Q4, with a total of 240 (T=240) observations. Since we wish to use a balanced panel data set, we excluded 36 variables because there were missing observations, and used a balanced panel for 210 variables. The variables are categorized into 14 groups [4]. 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. [5]

4.2 Estimation of factors

We adapted the methods proposed in Bai and Ng (2002) and Bai (2004) for choosing an ’optimal’ number of factors. The method proposed by Bai and Ng (2002), which is based on information criteria, led to the optimal number of factors being eight for the set of 100 sub-aggregate macroeconomic variables. To identify the optimal number of nonstationary factors, we applied the *iterated panel criterion* proposed by Bai (2004), and found four are $I(1)$ factors, and the other four are $I(0)$. We also 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 $I(0)$. Figure 1 shows time series plots of the estimated factors. The overall trends exhibited in these plots are consistent with the aforementioned observation that the factors $\{1, 2, 4, 5\}$ are $I(1)$ and the other four are $I(0)$. Augmented-Dickey Fuller (ADF) test suggest that the generated factors 1,2,4 and 5 are $I(1)$ and the other four are $I(0)$. Each of the estimated idiosyncratic error series $\{\hat{e}_{it}, t = 1, 2, \ldots\}$ was found to be $I(0)$.

For the data set in this empirical study, the panel data model and the forecasting model with a

[4] Details about the 14 categories can be found in the updated appendix of FRED QD data set $https : //s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QDappendix.pdf$.

[5] Transformation code for every series is given in the third row of the data set. FRED-MD listed seven different transformations to ensure the transformed series are $I(0)$.

mixture of stationary and nonstationary factors take the form,

$$X_{it} = \lambda_i' F_t + e_{it} = \lambda_i^{(1)\prime} E_t + \lambda_i^{(2)\prime} G_t + e_{it},$$

$$Y_{t+h} = \alpha' \tilde{E}_t + \beta' \tilde{G}_t + \omega_1 Y_t + \omega_2 Y_{t-1} + \omega_3 Y_{t-2} + \omega_4 Y_{t-3} + \epsilon_{t+h} \quad h > 0,$$

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.

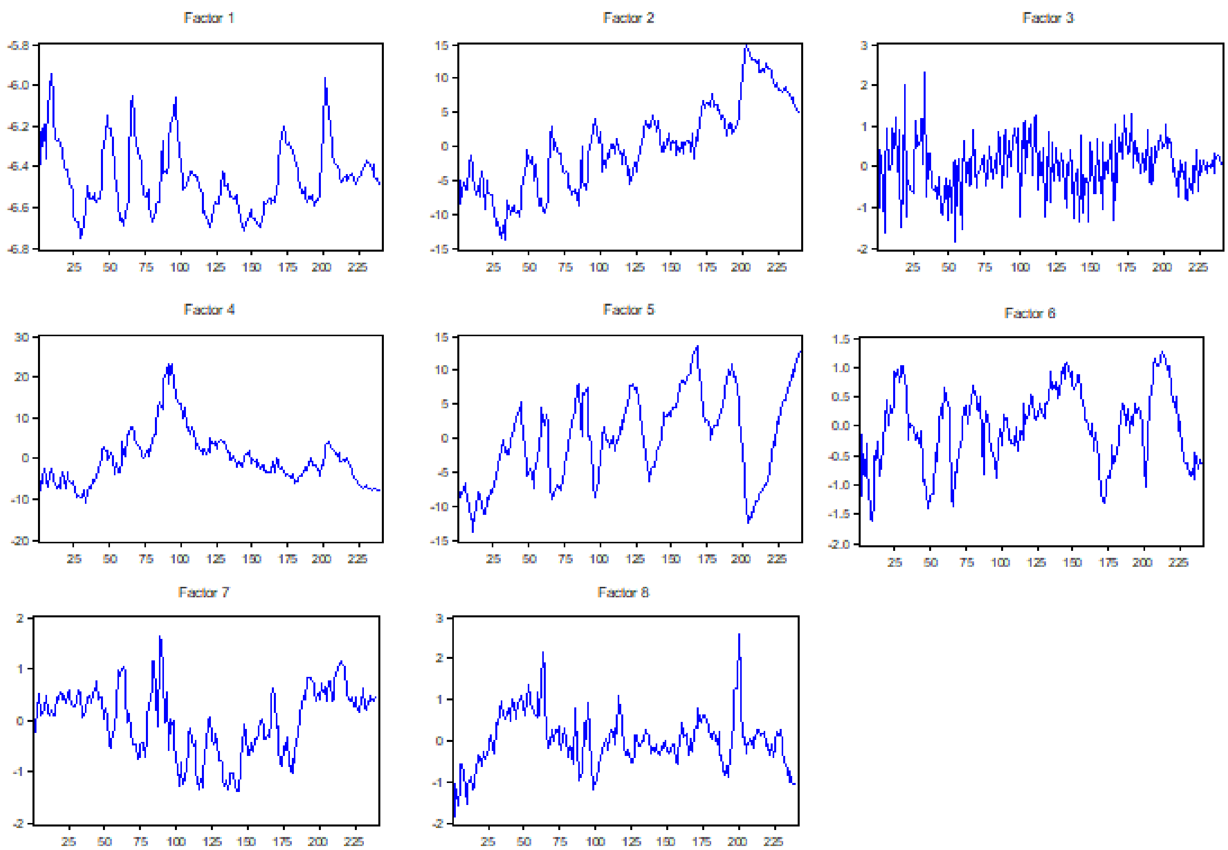

Figure 1: Generated factors from panel dataset of 100 variables

Plots of the two high-aggregate macroeconomic variables, GDP and IP, are presented in Figure 2, and plots of the eight generated factors are presented in Figure 1

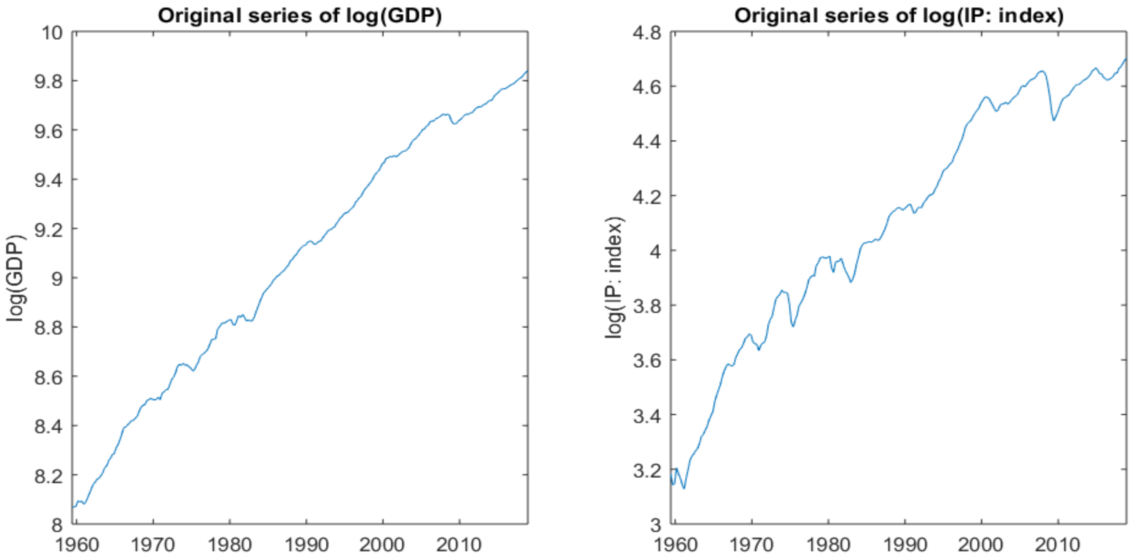

Figure 2: Time series plots of $\log(GDP)$ and $\log(IP)$ for 1959:Q1 – 2018:Q4.

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

In this study, we considered the following basic AR(4) and two other mixture-FAR models obtained by augmenting the mixture of eight $I(0)$ and $I(1)$ factors. The four models are specified as follows:

$$\text{Model 1}: Y_{t+h} = \alpha'\tilde{E}_t + \beta'\tilde{G}_t + \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h},$$

$$\text{Model 2}: Y_{t+h} = \alpha_1'\tilde{E}_t + \alpha_2'\tilde{E}_{t-1} + \beta_1'\tilde{G}_t + \beta_2'\tilde{G}_{t-1} + \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h},$$

$$\text{Model 3}: Y_{t+h} = \sum_{i=0}^{3}\alpha_{1+i}'\tilde{E}_{t-i} + \beta_1'\tilde{G}_t + \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h},$$

$$\text{Model 4}: Y_{t+h} = \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h}.$$

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 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. The theory developed in this paper is applicable to these models. In empirical studies such as the one in this section, an important requirement is to evaluate the adequacy of the proposed model. One standard way of checking adequacy of such specifications is by hypothesis testing. For forecasting, a particularly important criterion is out-of-sample forecast performance. Much of this section is based on evaluating such forecast performance.

The plots in Figure 3 and Figure 4 show the one-step ahead out-of-sample forecasts of $\log(GDP)$ and $\log(IP)$ for the period 2006:Q1 - 2018:Q4. Forecasts were made by expanding window, starting from the initial estimation period 1959:Q1 - 2005:Q4; this period has a window width 188 quarters, which we expect to be sufficiently large to apply the large sample theory developed in this paper. The plots indicate that the out-of-sample predictions of the two I(1) variables, GDP and IP, generally appear to be good since they are all close to the observed series.

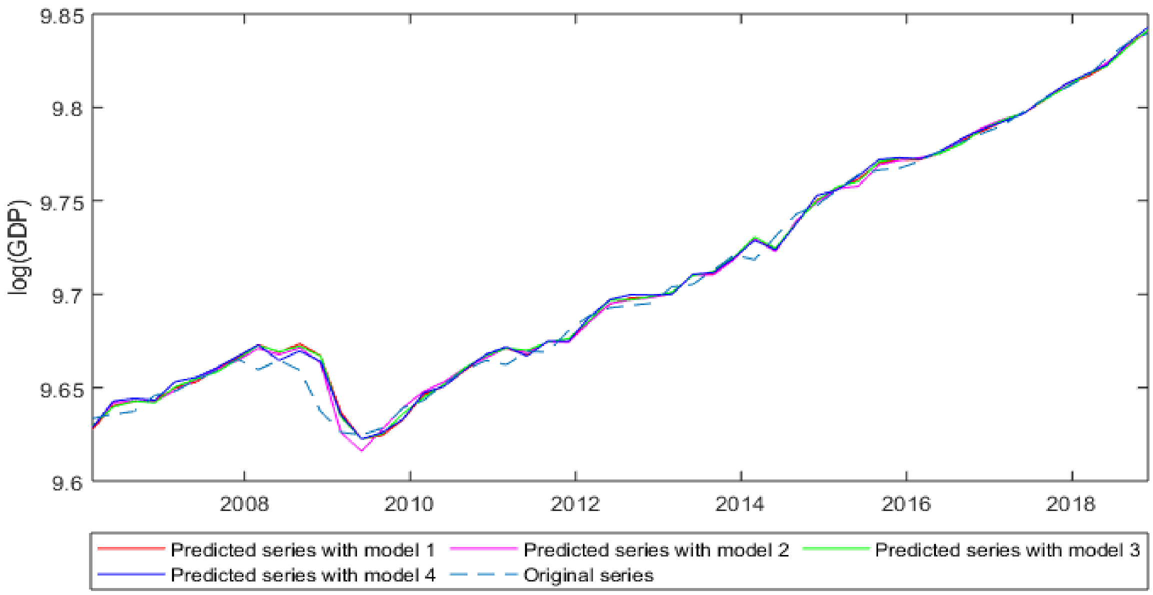

Figure 3: The observed $\log(GDP)$, and plots of one-step ahead out-of-sample forecasts of $\log(GDP)$ for 2006:Q1 – 2018:Q4.

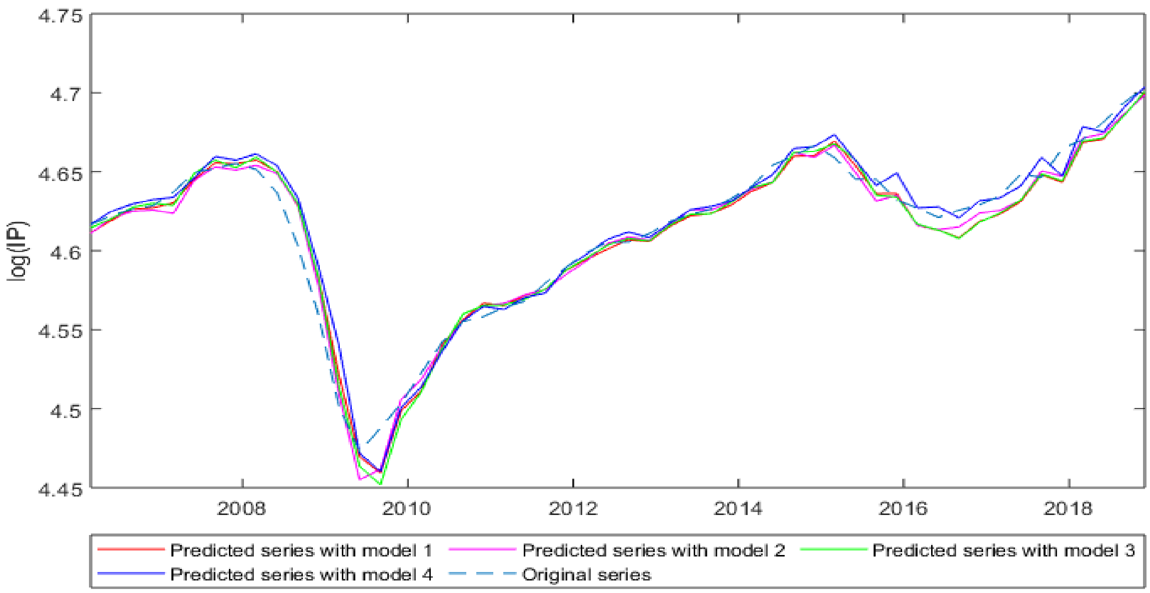

Figure 4: The observed $\log(IP)$, and plots of one-step ahead out-of-sample forecasts of $\log(IP)$ for 2006:Q1 – 2018:Q4.

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

We assess the relative predictive performance of the models in terms of the *out-of-sample* R^2, denoted R^2_{os}, defined in (24). In this subsection, AR(4) is used as the basic benchmark; later we consider nonstationary-FAR as the benchmark.

Table 2: First estimation and forecasting periods

First estimation period (T_1)	Out-of-sample period ($T - T_1$)	$(T - T_1)/T_1$
1959:Q1 - 1999:Q4	2000:Q1 - 2018:Q4	0.46
1959:Q1 - 2005:Q4	2006:Q1 - 2018:Q4	0.28
1959:Q1 - 2008:Q4	2009:Q1 - 2018:Q4	0.20

The three different out-of-sample periods.

We considered three different *first estimation periods.* [6] The R^2_{os} measures reported 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 outperforms the other two models.

[6]We also examined the rolling window forecast with 40 years for window size. Compared to the expanding window, there were no improvements in the forecast performance.

Table 3: Calculated R^2_{os} values for all three factor models compared to AR(4)

	Model 1		Model 2		Model 3	
	$\log(GDP)$	$\log(IP)$	$\log(GDP)$	$\log(IP)$	$\log(GDP)$	$\log(IP)$
First estimation period						
1959:Q1 - 1999:Q4	-0.0101	-0.1388	0.0681	0.1035	-0.0047	-0.2122
1959:Q1 - 2005:Q4	-0.0835	0.1729	0.1147	0.3258	-0.0119	0.1127
1959:Q1 - 2008:Q4	0.0929	0.0610	0.1581	0.2791	0.1820	-0.0139

The entries in the table are the calculated R^2_{os} values for one-step ahead expanding window out-of-sample forecasts of factor models relative to the AR(4) model.

4.3 Long term forecast evaluations

So far, we considered one-step ahead short-term forecasts. To evaluate the long term predictability of the models, we considered (expanding window) long-term forecasts with initial-sample period 1959:Q1 - 1998:Q4. For comparison purpose, we calculate R^2_{os} for different specifications of mixture-FAR model relative to AR(4). Figures 5 and 6 provide plots of R^2_{os} against the forecast horizon h.

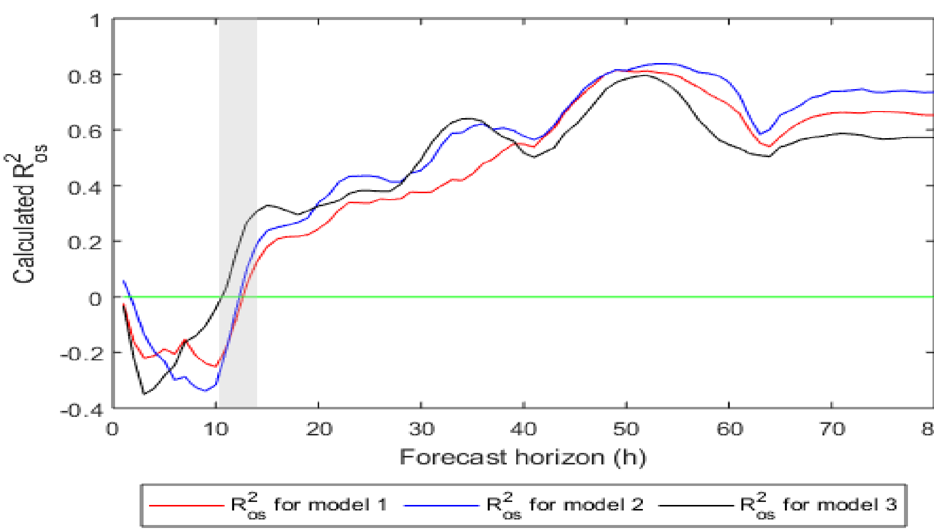

Figure 5: Performance of mixture-FAR relative to AR(4) for long-term forecast of $\log(GDP)$. The graph is a plot of R^2_{os} against the forecast horizon h

.

The plots in Figures 5 and 6 show that even for large forecast horizon ($h > 12$ here), the factor models perform better than the AR(4) model for forecasting GDP and IP. These results show that, in terms of R^2_{os} , the mixture-FAR model performs better than the AR(4) for short term and long term forecasting of GDP and IP.

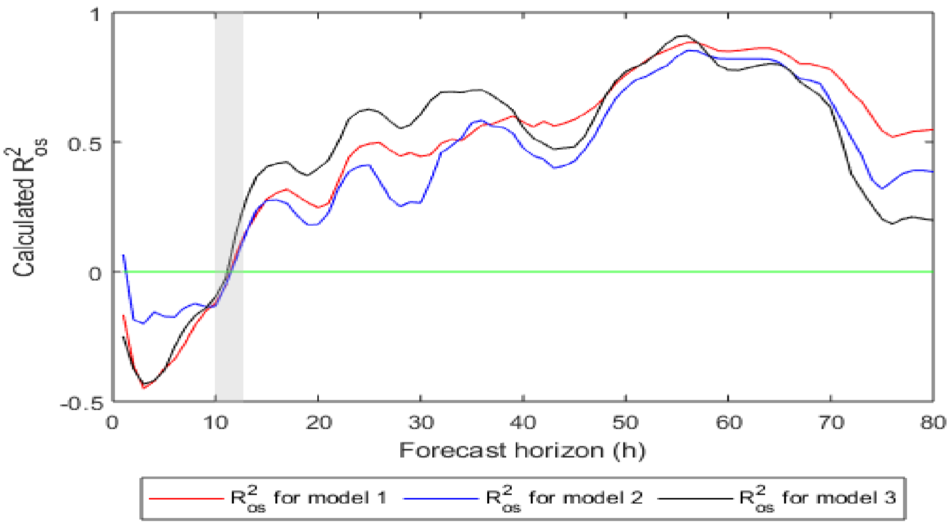

Figure 6: Performance of mixture-FAR relative to AR(4) for long-term forecast of $\log(IP)$. The graph is a plot of R^2_{os} against forecast horizon h.

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 as nonstationary-FAR model. The literature is not specific as to how we should ensure that the factors are nonstationary. To implement this method, we performed principal component analysis on X, and chose only the nonstationary factors for use as predictors in the prediction model (2). In this section, we compare the mixture-FAR and the aforementioned nonstationary-FAR methods for forecasting the nonstationary variables GDP and IP.

Next, we compare mixture-FAR with nonstationary-FAR for forecasting the nonstationary variables GDP and IP. Here, we compare the out-of-sample forecast performance using R^2_{os} for mixture-FAR relative to the corresponding nonstationary-FAR. We treat Model 1 as our mixture-FAR model. Therefore, we choose the corresponding nonstationary-FAR as follows with only the nonstationary factors:

$$\text{Nonstationary-FAR } (\textbf{Model 11}) : \; Y_{t+h} = \beta' \tilde{G}_t + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+h}.$$

The results are summarised in Table 4. Consider the entry 0.3078 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 30.08% lower for mixture-FAR compared to 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

Table 4: Calculated R^2_{os} for mixture FAR model compared to Model 11

First estimation period	$log(GDP)$	$log(IP)$
1959:Q1 - 1999:Q4	0.2103	0.1741
1959:Q1 - 2005:Q4	0.2121	0.3877
1959:Q1 - 2008:Q4	0.5004	0.3078

R^2_{os} for mixture-FAR model, Model 1, relative to the nonstationary-FAR model Model 11.

shows that mixture-FAR performed significantly better than nonstationary-FAR. for forecasting GDP and IP.

4.5 Prediction intervals

We computed the asymptotic theory-based and the bootstrap based 95% prediction intervals for one-step-ahead (recursive window) predictions for $\log(GDP)$ and $\log(IP)$. Assuming that the regression error (see ϵ_t in (2)) is normally distributed, we constructed the asymptotic theory-based point-wise 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 $\log(IP)$ to facilitate evaluation of the two prediction intervals in terms of out-of-sample coverage. Further, we also estimated the symmetric bootstrap t-percentile point-wise 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.

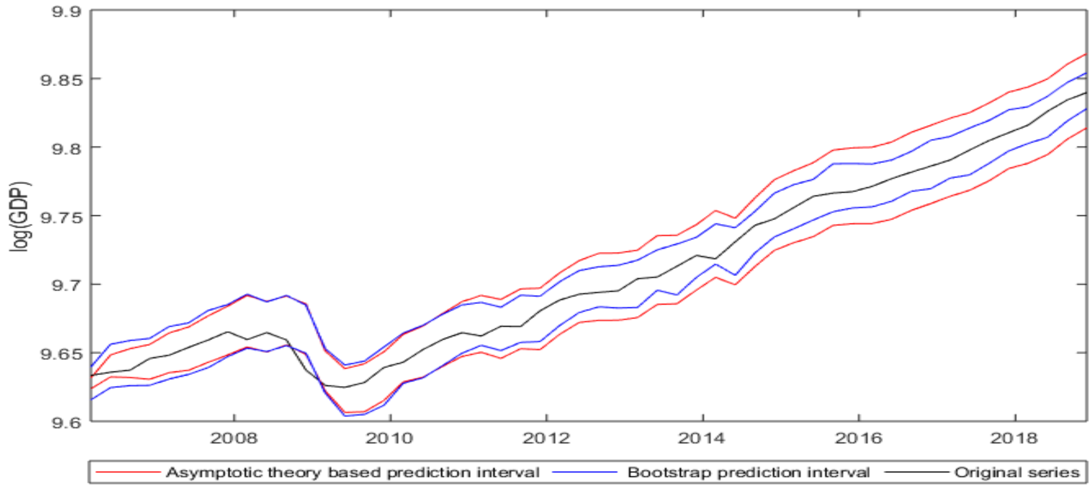

Figure 7: One-step ahead point-wise 95% prediction intervals for $\log(GDP)$ using mixture-FAR. One interval uses the asymptotic distribution of the forecast and the other uses residual bootstrap. These prediction intervals are for one-step ahead expanding window. The solid line in the middle is a plot of the observed values of $\log(GDP)$.

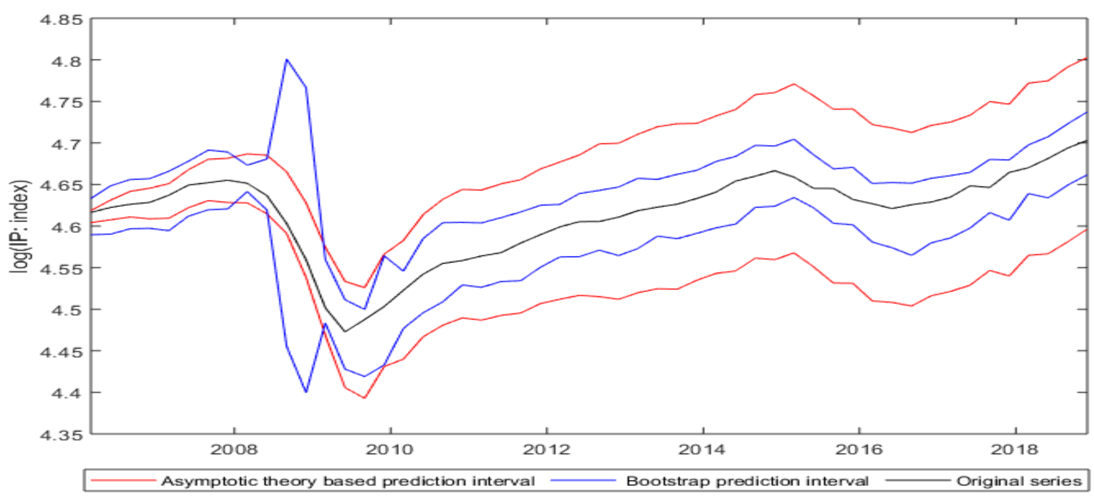

Figure 8: One-step ahead point-wise 95% prediction intervals for $\log(IP)$ using mixture-FAR. One interval uses the asymptotic distribution of the forecast and the other uses residual bootstrap. These prediction intervals are for one-step ahead expanding window. The solid line in the middle is a plot of the observed values of $\log(IP)$.

The 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. The Figure 8 shows that every observed value of IP lie 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 is 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.

5 Conclusion

This paper developed methodology for univariate time series forecasting of nonstationary macroeconomic variables, such as GDP and industrial production[IP], using high-dimensional panel data as predictors. To this end, we propose to replace the large number of potential predictors by a small number of factors estimated using the panel data, and use them as predictors in the forecasting model. The validity of this method for forecasting has been established when all variables are stationary (Bai and Ng (2006)), and also when they are all nonstationary (Choi (2017)), but not when the panel data consist of a mixture of stationary and nonstationary variables. In this setting, the panel data generate a mixture of stationary and nonstationary factors, and we use them for forecasting a nonstationary variable. To this end, this paper combines the ideas in the recent literature and develops the theory and methodology for constructing an asymptotically valid prediction interval. In a simulation study, we observed that the mixture-FAR developed in this paper performed significantly better than the one that uses only nonstationary variables. We applied the mixture-FAR model for forecasting GDP and IP. We assessed the out-of-sample forecast performance of the mixture-FAR relative to the corresponding (a) nonstationary-FAR and (b) the AR(4) model. 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.

Acknowledgements.

Sium Bodha Hannadige gratefully acknowledges the support from Monash University through a Postgraduate Scholarship. Jiti Gao gratefully acknowledges support from the Australian Government through the Australian Research Council under the Discovery Project scheme. Mervyn Silvapulle gratefully acknowledges support from the Australian Government through the Australian Research Council under the Discovery Project scheme.

Appendix A: Proofs

In this appendix we indicate the main steps for the proofs of Theorems 1 and 2, through a series of lemmas. The proofs of the lemmas will be available in Supplementary Materials to this paper, as a separate document. First, let us introduce the following notation:

$$\gamma_{st} = E\left(N^{-1}\sum_{i=1}^{N} e_{is}e_{it}\right), \quad \zeta_{st} = N^{-1}\sum_{i=1}^{N}\left(e_{is}e_{it} - E(e_{is}e_{it})\right), \tag{25}$$

$$\eta_{st} = N^{-1}F_s'\Lambda' e_t, \qquad \xi_{st} = N^{-1}F_t'\Lambda' e_s; \tag{26}$$

let us note that $\eta_{st} = \xi_{ts}$.

Lemma A. 1. *Suppose that Assumptions 1-3 are satisfied. For all N and T, we have*

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

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

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

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

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

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

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

Proof of Lemma 1 stated in the main paper is established the foregoing Lemma A.1, and Lemma A.2.

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

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

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

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

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

Lemma A. 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 $\delta_{NT} = min[\sqrt{N}, ||D_{1T}^{-1}||^{-1}]$.*

Lemma A. 5. *Suppose that Assumptions 1-6 are satisfied, and $T/N \to 0$. Then, as $T, N \to \infty$, we have $D_{1T}^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(\tilde{F}_t - HF_t\right)'(H^{-1})'\theta \xrightarrow{p} 0$.*

The proof of Theorem 1 uses the previous lemmas, in particular, Lemma A.4 and Lemma A.5.

The proof of Theorem 2 uses Theorem 1.

Appendix B: Simulation Results

In this section, we provide some simulation results on the inference together with the coverage rates for 95% prediction interval obtained from one-step ahead forecasts.

Table 5: 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	82	81	81	81	80	79	78	77	77	76	75	75
	50	84	82	82	80	80	79	79	79	78	76	77	76
	100	85	84	84	84	82	81	81	81	79	78	78	78
	200	86	85	83	83	82	82	80	80	79	79	77	77
DGP2	30	67	65	62	60	65	61	59	56	62	59	57	55
	50	72	69	66	66	69	66	64	63	66	64	62	60
	100	77	75	73	72	74	71	70	68	71	68	66	66
	200	79	76	75	73	75	73	71	70	73	70	68	68
DGP3	30	68	65	62	61	64	61	59	58	62	58	57	56
	50	73	69	67	66	70	66	64	62	67	64	61	61
	100	77	75	72	72	74	71	70	68	71	68	67	65
	200	78	76	74	73	75	73	71	70	72	71	69	67

The error distribution of the forecasting model is normal.

Tables 6 and 8 provide the simulation results on bias of OLS estimates and coverage rates of 95% confidence intervals.

Table 6: Bias of OLS estimates and coverage rates of 95% confidence intervals for parameters in the regression model when the two factors are independent

		α				β				ω			
DGP	N\T	30	50	100	200	30	50	100	200	30	50	100	200
						bias ($\times 10^{-4}$)							
DGP1	30	-13	-65	35	71	-27	-158	17	-186	-250	-129	-40	-5
	50	-14	-20	-48	33	-107	-298	-87	-129	-263	-142	-51	-13
	100	8	33	26	-13	-31	-171	37	-46	-283	-149	-77	-35
	200	-23	-45	-11	0	-150	-281	96	-32	-318	-165	-83	-40
DGP2	30	38	-58	35	79	21	-166	26	-186	-39	-5	19	38
	50	-2	-23	-33	37	-100	-307	-68	-120	-71	-26	3	13
	100	6	37	24	-9	-40	-176	33	-47	-87	-41	-13	-5
	200	-33	-18	-12	3	-158	-252	109	-35	-92	-46	-25	-9
DGP3	30	30	-77	14	90	-123	-140	116	-106	-44	-2	30	45
	50	29	-23	-39	46	-113	-118	-46	-112	-60	-35	1	25
	100	41	12	28	-20	-33	-285	126	4	-71	-43	-26	1
	200	1	9	-18	9	-189	-312	99	11	-101	-65	-19	-9
						coverage rate (%)							
DGP1	30	88	95	99	99	69	70	74	76	99	99	99	99
	50	88	95	99	99	68	70	71	72	99	99	99	99
	100	87	95	98	99	66	68	69	70	99	99	99	99
	200	87	94	98	99	65	68	68	69	99	99	99	99
DGP2	30	84	96	97	98	64	73	82	87	99	99	99	99
	50	81	93	97	98	66	74	82	86	99	99	99	99
	100	78	92	97	98	64	73	82	86	99	99	99	99
	200	77	92	97	97	64	74	81	87	99	99	99	99
DGP3	30	84	95	97	98	65	74	83	87	98	99	99	99
	50	82	93	97	98	65	73	82	86	98	98	99	99
	100	78	92	97	98	64	73	82	86	97	98	99	99
	200	76	91	96	97	65	74	86	87	96	99	99	99

The true parameter values of (α, β, ω) are $(0.5, 1.0, 0.5)$. Top panel contains the bias of OLS estimates of the regression parameters in the regression model. The number of simulations is 5000. The two coefficients corresponding to the generated two factors are α and β respectively, and ω is the coefficient of the time lag of Y.

Table 7: Bias of OLS estimates and coverage rates of 95% confidence interval for parameter estimates when the two factors are correlated with $\rho = 0.5$

		α				β				ω			
	N\T	30	50	100	200	30	50	100	200	30	50	100	200
						bias ($\times 10^{-4}$)							
DGP1	30	-21	-72	38	80	-154	-43	-40	-184	-376	-220	-104	-60
	50	-49	-28	-50	35	-41	-369	18	-99	-360	-213	-101	-50
	100	-40	44	19	-10	-46	-141	58	-52	-372	-208	-117	-59
	200	-30	-55	-6	-2	26	-122	32	-28	-406	-209	-116	-56
DGP2	30	16	-55	35	86	-12	-52	-44	-190	-136	-84	-47	-30
	50	-44	-29	-48	43	-48	-373	23	-92	-154	-87	-44	-28
	100	11	22	22	-10	-61	-145	78	-65	-127	-80	-49	-30
	200	-29	-32	-8	3	7	-102	50	-37	-142	-77	-50	-23
DGP3	30	23	-79	31	95	-1	-34	40	-95	-147	-86	-40	-22
	50	-19	-15	-31	44	-78	-86	-8	-170	-126	-88	-46	-23
	100	-13	31	31	-6	159	-192	105	10	-109	-86	-60	-29
	200	-7	6	-24	11	-46	-126	40	-36	-139	-87	-43	-22
						coverage rate (%)							
DGP1	30	87	94	99	99	70	72	76	77	99	99	99	99
	50	88	94	99	99	70	71	72	74	99	99	99	99
	100	87	95	98	99	68	70	70	72	99	99	99	99
	200	87	94	98	99	70	69	68	70	99	99	99	99
DGP2	30	83	94	97	98	64	73	82	87	99	99	99	99
	50	80	93	97	98	65	75	81	87	99	99	99	99
	100	78	92	97	98	64	73	82	87	99	99	99	98
	200	77	92	96	98	63	74	81	86	99	99	98	99
DGP3	30	84	95	97	98	63	74	81	86	99	99	99	99
	50	81	93	97	98	63	74	81	86	99	99	99	99
	100	78	92	97	98	63	75	82	86	99	99	99	99
	200	77	91	97	98	64	73	81	86	99	99	99	99

The true parameter values of (α, β, ω) are $(0.5, 1.0, 0.5)$. Top panel contains the bias of OLS estimates of the parameters in the regression model. The number of simulations is 5000. The two coefficients corresponding to the generated factors are α and β respectively, and ω is the coefficient of the time lag of Y.

Table 8: Bias of OLS estimates and coverage rates of 95% confidence intervals for parameter estimates when the two factors are correlated with $\rho = 0.9$

		α				β				ω			
	N\T	30	50	100	200	30	50	100	200	30	50	100	200
						bias ($\times 10^{-4}$)							
DGP1	30	-14	-81	55	86	339	-26	28	-139	-624	-409	-214	-152
	50	-73	-14	-44	32	198	-33	75	-93	-593	-358	-201	-114
	100	-52	60	9	3	292	27	217	-100	-582	-353	-196	-105
	200	-37	-67	-4	-2	147	89	211	-14	-610	-323	-185	-89
DGP2	30	17	-62	49	94	305	-15	37	-130	-398	-305	-213	-175
	50	-78	-27	-47	42	260	-52	64	-92	-349	-241	-155	-115
	100	15	10	5	0	312	35	217	-82	-290	-209	-145	-92
	200	-50	-40	-14	7	146	111	211	-9	-295	-184	-114	-60
DGP3	30	33	-68	26	105	195	45	-79	-14	-408	-307	-218	-167
	50	-105	-30	-27	33	126	13	-54	52	-362	-241	-175	-127
	100	-31	35	31	8	453	124	44	-54	-283	-201	-132	-81
	200	-37	-3	-22	9	398	72	66	-35	-283	-196	-109	-64
						coverage rate (%)							
DGP1	30	86	93	99	99	73	75	77	78	99	99	99	99
	50	87	94	99	99	72	73	74	74	99	99	99	99
	100	85	94	98	99	71	72	72	72	99	99	99	99
	200	85	94	99	99	70	72	72	72	99	99	99	99
DGP2	30	82	95	98	98	64	75	82	87	99	99	99	99
	50	80	93	97	98	64	74	82	87	99	99	99	99
	100	78	92	97	98	65	74	81	86	99	99	99	99
	200	76	91	97	98	64	75	82	87	99	99	99	99
DGP3	30	82	94	97	98	63	74	81	85	99	99	99	99
	50	80	93	97	98	64	74	81	86	99	99	99	99
	100	78	92	97	98	64	75	82	86	99	99	99	99
	200	77	92	97	98	66	74	82	86	99	99	99	99

The true parameter values (α, β, ω) are $(0.5, 1.0, 0.5)$. Top panel contains the bias of the OLS estimates of the parameters in the regression model. The number of simulations is 5000. the two coefficients corresponding to the generated factors are α and β respectively, and ω is the coefficient of the time lag of Y.

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. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and 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., and 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. and 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. and Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, 182(1):156–173.

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

Moon, H. R. and 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. and Wijler, E. (2019). High-dimensional forecasting in the presence of unit roots and cointegration. Working paper, Department of Quantitative Economics, Maastricht University.

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

Stock, J. H. and Watson, M. W. (1998b). Diffusion indexes. Technical report, National Bureau of Economic Research.

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

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