Differencing versus Non-Differencing in Factor-Based Forecasting*

In Choi† and Hanbat Jeong‡

First Draft: July, 2018; First revision: July, 2019; This draft: January, 2020

Abstract

This paper studies performance of factor-based forecasts using differenced and non-differenced data. Approximate variances of forecasting errors from the two forecasts are derived and compared. It is reported that the forecast using non-differenced data tends to be more accurate than that using differenced data. This paper conducts simulations to compare root mean squared forecasting errors (RMSFE) of the two competing forecasts. Simulation results indicate that forecasting using non-differenced data performs better. The advantage of using non-differenced data is more pronounced when a forecasting horizon is long and the number of factors is large. This paper applies the two competing forecasting methods to 68 $I(1)$ monthly U.S. macroeconomic variables across a range of forecasting horizons and sampling periods. We also provide detailed forecasting analysis on U.S. inflation and Industrial Production. We find that forecasts using non-differenced data tend to outperform those using differenced data.

Keywords: Factor-based forecasting, Nonstationary factors, $I(1)$ monthly U.S. macroeconomic variables, U.S. inflation rates, Industrial Production

JEL classification: C38, C53

*We thank Michael W. McCracken (Co-editor) and two anonymous referees for their constructive comments that substantially improved our paper.

†Corresponding author. Department of Economics, Sogang University, 347 Baekbeom-ro, Mapo-gu, Seoul, Korea. E-mail: inchoi@google.com, inchoi@sogang.ac.kr. Personal Webpage: http://inchoi.sogang.ac.kr.

‡Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. E-mail: jeong.181@buckeyemail.osu.edu

1 Introduction

Factor models have been utilized for various purposes in economics and finance: (i) forecasting, (ii) formulation of economic indicators, (iii) policy analysis, (iv) instrumental variables estimation, and (v) modeling cross-sectional correlation. See Breitung and Eickmeier (2006), Bai and Ng (2008), Stock and Watson (2011) and Breitung and Choi (2013) for surveys on the related literature. Among many applications of factor models, this paper focuses on factor-based forecasting. The factor-based prediction was initiated by Stock and Watson (1999, 2002) and Bai and Ng (2006), and has been used profitably for forecasting important economic variables. Recent related studies for this issue are Banerjee, Marcellino, and Masten (2014), Gonçalves and Perron (2014), Cheng and Hansen (2015), and Gonçalves, McCracken, and Perron (2017). It extracts estimated factors from a large dataset and employs them in forecasting regressions along with other observed regressors. It has become one of the standard methods of forecasting for an environment with a large dataset. So far, following Stock and Watson (1999, 2002) and Bai and Ng (2006), it has been a common practice to use differenced data in factor-based forecasting because differenced data are deemed to be stationary and the standard theory of inference can be applied for differenced data. However, it is uncertain whether the common practice provides best forecasting performance.[1] Hence, the purpose of this paper is to study whether differenced data offer better forecasting performance than non-differenced (or level) data when the factor-based forecasting model is used. In our framework, estimated factors from non-differenced data are nonstationary. Properties of the estimators of nonstationary factors and related model selection criteria are studied in Bai (2004). In addition, Choi (2017) considers the generalized principal component estimator for nonstationary factors and its asymptotic properties.

This paper's contributions are threefold. First, we derive and compare approximate variances of forecasting errors from the regression in levels and that in differences. The factor-based prediction using level data employs only $I(1)$ variables for estimating factor spaces. In contrast to that forecasting method, the forecasting method based on differenced data employs $I(0)$ variables in addition to differenced $I(1)$ variables. The derived approximate variances of forecasting errors reveal that the forecasts from the regression in levels tend to be more accurate than those from the regression in differences. We also find that the approximate variance of forecasts based on differenced data increases as does a forecasting horizon. By

[1] In building forecasting models, Diebold and Kilian (2000) suggests that routinely differencing for stationarity is not recommended. Even though their study is not a factor-based forecasting model, their point shares the same critical mind with us.

contrast, the approximate variance of forecasts based on level data does not vary with a forecasting horizon.

Second, we compare root mean squared forecasting errors (RMSFE) of the two forecasts via Monte Carlo simulations. Simulation results indicate that the forecasting method using level data usually performs better. Even though we throw out a lot of stationary variables for factor estimation, the level-based forecasts outperform those using differenced data in most cases. The advantage of using level data is more pronounced when a forecasting horizon is long and the number of factors is large. A practical implication of these results is that level-based forecasting should be preferred for the long-run prediction. Using differenced data can be preferred if a researcher can employ a lot of $I(0)$ variables for estimating factor spaces and a forecasting horizon is short, or if forecasting errors are highly persistent.

Third, we apply the two competing forecasts, one using differenced data and the other using non-differenced ones, to various $I(1)$ monthly U.S. macroeconomic variables. The dataset is borrowed from McCracken and Ng (2016), which is a newer version of Stock and Watson's (2005) dataset. We compare their performance based on ex-post RMSFE. Instead of reporting all results, we provide key summary measures showing the relative efficiency of the level-based forecasts. We make a comparison between the two forecasting methods across various forecasting horizons and three sampling periods: (i) the pre-Great Moderation period (1960 - 1983), (ii) the Great Moderation period (1984 - June 2007), and (iii) the crisis and aftermath period (July 2007 - October 2018). We detect heterogeneous patterns of relative efficiency measures across the sampling periods as well as forecasting horizons. We find that forecasts using non-differenced data tend to outperform those using differenced data, in particular, in the Great Moderation and the crisis and aftermath periods. As contrary to the forecasting method based on differenced data, the forecasting method using level data shows performance robust to forecasting horizons. It verifies our theoretical and simulation results that our level-based forecasting approach is better suited for the long-run prediction.

We also provide detailed forecasting analysis on the CPI-based U.S. inflation rate and the U.S. Industrial Production (IP). For the two variables, we report ex-post RMSFEs, the Diebold-Mariano's (1995) test results, and persistency measures of forecasting errors. For the U.S. inflation rate and the IP index, the relative efficiency of the level-based forecasting method is most pronounced in the crisis and aftermath period. The efficiency gain of our level-based method for the U.S. inflation rate is smallest in the Great Moderation period. For the IP index, we observe the smallest efficiency gain of our method in the pre-Great Moderation period. Last, estimated forecasting errors in the IP index are more persistent than those in the

U.S. inflation rate.

Notation: The following notation will be used throughout this paper. For arbitrary matrices X and Y, $X \oplus Y = \begin{bmatrix} X & \mathbf{0} \\ \mathbf{0} & Y \end{bmatrix}$. For a $k \times 1$ vector $a = (a_1, \cdots, a_k)'$, $\|a\| = \sqrt{a_1^2 + \cdots + a_k^2}$ denotes the Euclidean norm of a. All the limits are taken as $N, T \to \infty$. Convergence in probability and convergence in distribution are denoted as $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$, respectively. The usual difference operator is denoted as Δ (i.e., $\Delta x_t = x_t - x_{t-1}$). $[z]$ denotes the integer part of z.

2 Factor-based forecasting

This section introduces two approaches of factor-based forecasting and studies their theoretical properties. One method employs non-differenced (level) data, and the other uses differenced data. The former approach is new and has seldom been used in the literature. The latter approach transforms all the data such that they become stationary. It has been used widely in practice, and its theoretical aspects are discussed in Bai and Ng (2006).

2.1 Forecasting using non-differenced data

Consider forecasting an $I(1)$ variable y_{T+h} in a data-rich environment, where T is the size of available sample and h is a forecasting horizon. The factor-augmented forecasting equation using non-differenced (level) data is written as

$$y_{t+h} = \mu + \alpha' F_t + \beta' Z_t + \epsilon_{t+h}, \; (t = 1, \cdots, T-h), \tag{1}$$

where μ is a constant, F_t is a vector of $I(1)$ factors from model (2) below, Z_t is a $K \times 1$ vector of observable variables that may include a linear time trend, $I(1)$ and $I(0)$ variables, and ϵ_{t+h} is an $I(0)$ error component. If ϵ_{t+h} is $I(1)$, the case of spurious regression occurs. We forecast y_{T+h} in two steps. First, the common factors F_t are extracted from $I(1)$ predictors X_t, an $N_1 \times 1$ vector. Second, we estimate the coefficients μ, α and β by replacing F_t with its estimator from the first step and forecast y_{T+h} in the usual manner.

Now we discuss how to estimate nonstationary factor spaces from the predictors X_t. Suppose that the variables X_t are modeled as

$$X_t = \Lambda F_t + e_t, \; (t = 1, \cdots T), \tag{2}$$

where $\Lambda = [\lambda_1, \cdots, \lambda_{N_1}]'$ is an $N_1 \times r$ matrix of factor-loadings, F_t is an $r \times 1$ vector of latent factors and $e_t = [e_{1t}, ..., e_{N_1 t}]'$ is a vector of stationary, idiosyncratic components of the model.

Assume that $\{F_t\}$ is a nonstationary process represented by

$$F_t = F_{t-1} + u_t,$$

where $\{u_t\}$ is a vector of zero-mean, weakly stationary process and F_0 is a random vector. It is assumed that there is no cointegration relationship among the elements of F_t. The number of factors r is assumed to be known in Sections 2 - 3. Methods to estimate r are introduced in Bai (2004). Let N_0 denote the number of available $I(0)$ variables so that the total number of predictors becomes $N = N_0 + N_1$. We assume that $\frac{N_1}{N} \to c$ as $N \to \infty$, where $c \in (0,1)$. In addition, the number of factors r is assumed to be unrelated to N_0, N_1 and N.

Estimators of Λ and F_t can be obtained by the principal component estimation method. This estimation method is based on solving the quadratic optimization problem:

$$\min_{\Lambda, F_1, ..., F_T} \frac{1}{N_1 T} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t).$$

Let $F = [F_1, ..., F_T]'$ and $X = [X_1, \cdots, X_T]'$. With the standardization $F'F = T^2 \times I_r$, the principal component estimator (PCE) of the factor space of F, denoted by $\tilde{F}$, is T^2 times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of XX'. The PCE of Λ is given by $\tilde{\Lambda}^L = \frac{1}{T^2} X' \tilde{F}$.

Under Assumptions A - D given in Bai (2004)[2], we have

$$\min\left\{N_1, T^2\right\} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{F}_t - H_L' F_t \right\|^2 \right) = O_p(1), \tag{3}$$

where H_L is an $r \times r$ rotation matrix that corresponds to H_2 in Bai's (2004) Corollary 1. Also, if $\frac{N_1}{T^3} \to 0$ and Assumptions E - G of Bai (2004) are additionally assumed, for each t,

$$\sqrt{N_1} \left(\tilde{F}_t - H_L' F_t \right) \xrightarrow{d} \left(V^L\right)^{-1} Q^L \; N(0, \Sigma_t), \tag{4}$$

[2] For details, see the first section of the supplement to this paper.

where V^L and Q^L are $r \times r$ matrices[3], and

$$\Sigma_t = \lim_{N_1 \to \infty} \frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} E(\lambda_i \lambda_j' e_{it} e_{jt}).$$

Note that Σ_t is the asymptotic variance of $\frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} \lambda_i e_{it}$. The asymptotic results (3) and (4) are essential for proving Theorem 3 in Subsection 2.3.

Once $\{\tilde{F}_t\}$ is available, we replace F_t with $\tilde{F}_t$ in model (1). Then, the forecasting equation with $\tilde{F}_t$ is written as

$$\begin{aligned} y_{t+h} &= \mu + \alpha' H_L'^{-1} \tilde{F}_t + \beta' Z_t + \epsilon_{t+h} + \alpha' H_L'^{-1} \left(H_L' F_t - \tilde{F}_t \right) \\ &= \delta^{L\prime} \tilde{L}_t + \epsilon_{t+h} + \alpha' H_L'^{-1} \left(H_L' F_t - \tilde{F}_t \right), \end{aligned} \tag{5}$$

where $\delta^L = \left(\mu, \alpha' H_L'^{-1}, \beta'\right)'$ and $\tilde{L}_t = \left(1, \tilde{F}_t', Z_t'\right)'$. Model (5) provides the OLS estimator $\hat{\delta}^L$, and the forecast of y_{T+h} is $\hat{y}^L_{T+h|T} = \hat{\delta}^{L\prime} \tilde{L}_T$.

The forecasting error of $\hat{y}^L_{T+h|T}$ is written as

$$\hat{y}^L_{T+h|T} - y_{T+h} = \left(\hat{\delta}^L - \delta^L\right)' \tilde{L}_T + \alpha' H_L'^{-1} \left(\tilde{F}_T - H_L' F_T\right) - \epsilon_{T+h}. \tag{6}$$

Observe that the forecasting error involves three terms which bring sampling variability: (i) $\hat{\delta}^L - \delta^L$, (ii) $\tilde{F}_T - H_L' F_T$, and (iii) ϵ_{T+h}. Thus, to obtain the asymptotic variance of $\hat{y}^L_{T+h|T} - y_{T+h}$, we need to derive the asymptotic distribution of $\hat{\delta}^L - \delta^L$ as an initial step. To this ends, we make the following assumption.

Assumption 1 *Let $L_t = (1, F_t', Z_t')'$. For $h = 1, 2, \cdots$, the following holds.* $\begin{pmatrix} D_T^{-1} \sum_{t=1}^{T-h} L_t L_t' D_T^{-1} \\ D_T^{-1} \sum_{t=1}^{T-h} L_t \epsilon_{t+h} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \Sigma_L \\ \Sigma_{\epsilon L}^{1/2} N(0, I) \end{pmatrix}$ *where D_T is a diagonal matrix whose elements are functions of T making $D_T^{-1} \sum_{t=1}^{T-h} L_t L_t' D_T^{-1} = O_p(1)$, $\Sigma_L > 0$ and $\Sigma_{\epsilon L} > 0$ with probability one, and $\Sigma_{\epsilon L}^{1/2}$ is independent of $N(0, I)$.*

Assumption 1 imposes conditions on regressors $\{L_t\}$ and forecasting errors $\{\epsilon_t\}$ in the forecasting equation. This is a counterpart of Bai and Ng's (2006) Assumption E, and we can allow various scenarios for $\{L_t\}$.

[3]Note that $H_L = \left(\frac{\Lambda'\Lambda}{N_1}\right)\left(\frac{F'\tilde{F}}{T^2}\right)\left(V^L_{N_1 T}\right)^{-1}$ where $V^L_{N_1 T} = diag\left(v^L_{N_1 T,1}, \cdots, v^L_{N_1 T,r}\right)$, with $v^L_{N_1 T,1} \geq \cdots \geq v^L_{N_1 T,r} > 0$, and $v^L_{N_1 T,j}$ denotes the jth-eigenvalue of $\frac{XX'}{N_1 T^2}$. The limit distribution of $V^L_{N_1 T}$ is $V^L = diag\left(v^L_1, \cdots, v^L_r\right)$, where $v^L_1 > \cdots > v^L_r > 0$ and v^L_j denotes the j^{th}-eigenvalue of $\Sigma_\Lambda^{1/2} \int_0^1 B_F(s) B_F'(s) ds \Sigma_\Lambda^{1/2}$. Υ^L is the corresponding eigenvector matrix such that $\Upsilon^{L\prime}\Upsilon^L = I_r$. Then, $Q^L = \left(V^L\right)^{1/2} \Upsilon^{L\prime} \Sigma_\Lambda^{-1/2}$.

Assumption 1 can cover various cases for which $D_T^{-1}\sum_{t=1}^{T-h} L_t L_t' D_T^{-1}$ is well defined in the limit. If $\{Z_t\}$ is a zero-mean, stationary process,

$$\Sigma_L = \begin{bmatrix} 1 & \int_0^1 B_F'(s)ds \\ \int_0^1 B_F(s)ds & \int_0^1 B_F(s)B_F'(s)ds \end{bmatrix} \oplus \Sigma_z,$$

where $B_F(s)$ is the weak limit of $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} u_t$ $(0 \leq s \leq 1)$, Σ_z $=\text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'$ with $D_T = diag[\sqrt{T}, T, ..., T, \sqrt{T}, ..., \sqrt{T}]$. If $\{Z_t\}$ is an $I(1)$ process, choosing $D_T = diag[\sqrt{T}, T, ..., T]$ yields

$$\Sigma_L = \begin{bmatrix} 1 & \int_0^1 B_{FZ}'(s)ds \\ \int_0^1 B_{FZ}(s)ds & \int_0^1 B_{FZ}(s)B_{FZ}'(s)ds \end{bmatrix},$$

where $B_{FZ}(s) = (B_F'(s), B_Z'(s))'$ and $B_Z(s)$ is the weak limit of $\frac{1}{\sqrt{T}}Z_{[Ts]}$.

Assumption 1 also assumes the limiting distribution of $D_T^{-1}\sum_{t=1}^{T} L_t\epsilon_{t+h}$. If (i) $\{F_t\}$ is independent of $\{\epsilon_t\}$ and (ii) $\{Z_t,\ \epsilon_{t+h}\}$ is stationary and ergodic, we have $D_T = diag[\sqrt{T}, T, ..., T, \sqrt{T}, ..., \sqrt{T}]$ with

$$\Sigma_{\epsilon L} = \sigma_\epsilon^2 \begin{bmatrix} 1 & \int_0^1 B_F'(s)ds \\ \int_0^1 B_F(s)ds & \int_0^1 B_F(s)B_F'(s)ds \end{bmatrix} \oplus \lim_{T\to\infty}\frac{1}{T}E(\sum_{t=1}^{T-h} Z_t\epsilon_{t+h})(\sum_{t=1}^{T-h} Z_t\epsilon_{t+h})',$$

where σ_ϵ^2 is the long-run variance of $\{\epsilon_t\}$. If (i) $\{F_t\}$ is independent of $\{\epsilon_{t+h}\}$ and (ii) $\{Z_t\}$ is $I(1)$ and independent of $\{\epsilon_{t+h}\}$, we have

$$\Sigma_{\epsilon L} = \sigma_\epsilon^2 \begin{bmatrix} 1 & \int_0^1 B_{FZ}'(s)ds \\ \int_0^1 B_{FZ}(s)ds & \int_0^1 B_{FZ}(s)B_{FZ}'(s)ds \end{bmatrix},$$

with $D_T = diag[\sqrt{T}, T, ..., T]$. Other cases of Z_t containing nonstochastic trend functions can also be considered in a similar manner.

The asymptotic distribution of the OLS estimator $\hat{\delta}^L$ is reported in the following lemma. This lemma will be used to derive the approximate variance of forecasting errors in Theorem 3.

Lemma 1 *Suppose that assumptions for the asymptotic result (3) and Assumption 1 hold. In addition, assume for $M > 0$*

$$\text{(i)} \sup_{N_1,T}\sum_{s=1}^{T}\frac{1}{N_1}\left|E\left(e_s' e_t\right)\right| < M \text{ for each } t$$

and

$$\text{(ii)}\ Var\left(\frac{1}{\sqrt{N_1}}E_T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N_1} Z_t e_{it}\lambda_i'\right) = O\left(1\right),$$

where E_T is a diagonal matrix of dimension K making $Var\left(E_T^{-1}\sum_{t=1}^{T} Z_t\right) = O(1)$. Let $\Sigma_\Lambda = p\lim_{N_1\to\infty}\frac{\Lambda'\Lambda}{N_1}$ and $\Psi_L = 1 \oplus \left(V^L\right)^{-1/2} \Upsilon_L' \Sigma_\Lambda^{1/2} \oplus I_K$ (see footnote 3 for the definitions of V^L and Υ_L). Then,

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

If $\{F_t, Z_t\}$ and $\{\epsilon_t\}$ are independent, $\Psi_L'^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}^{1/2}$ is independent of $N(0, I)$.

The additional assumptions (i) and (ii) are introduced to prove that the errors stemming from replacing F_t with $\tilde{F}_t$ in model (1) are negligible in the limit. More specifically, they are required to prove $D_T^{-1}\sum_{t=1}^{T} L_t\left(\tilde{F}_t - H_L' F_t\right)' = o_p(1)$. Assumption (i) is similar to Assumption E of Bai (2004), and characterizes weak dependence of idiosyncratic components across t. Assumption (ii) imposes a restriction on dependence among $\{Z_t\}$, $\{e_{it}\}$ and $\{\lambda_i\}$. For a stationary factor-augmented regression model, a similar assumption can be found in Assumption 3 (c) of Gonçalves and Perron (2014). Besides the assumptions given Lemma 1, the asymptotic result (3) also plays a crucial role in proving it. See the proof of Lemma 1 for its details.

In the conventional differencing-based method, we need $\frac{\sqrt{T}}{N} \to 0$ to eliminate the asymptotic bias of the OLS estimator for the forecasting equation that stems from factor estimation. As Ludvigson and Ng (2011) and Gonçalves and Perron (2014) point out, the asymptotic bias appears in the OLS estimator if $\frac{\sqrt{T}}{N} \to c_0 > 0$. According to Lemma 1, however, restrictions on the ratio of N_1 and T are not required to eliminate the asymptotic bias of $\hat{\delta}^L$ that originates from factor estimation. An intuitive reason for the difference between the nonstationary and stationary cases is the strong signal from $I(1)$ regressors. In a similar vein, OLS estimators from cointegrating regressions are consistent and free of asymptotic bias even in the presence of nonzero covariance between the regressors and errors. Because technical explanations for the difference are lengthy and involve many, peripheral notations, they are relegated to the supplement to this paper.

2.2 Forecasting using differenced data

Alternatively, we can forecast y_{T+h} using the conventional, differencing-based approach. Letting $w_{t+h} = \Delta y_{t+h}$, the forecasting regression equation based on this approach is

$$\begin{aligned} w_{t+h} &= \mu + \alpha'\Delta F_t + \beta'\Delta Z_t + \Delta\epsilon_{t+h},\ (t = 2, \cdots, T-h) \quad (7)\\ &= \mu + \alpha' u_t + \beta'\Delta Z_t + \Delta\epsilon_{t+h}. \end{aligned}$$

If equation (1) represents the true data generating process, the constant term in equation (7) should be absent. But it will be retained here following the convention of a linear regression.

For the first stage of forecasting, we difference model (2) and combine the resulting model with stationary variables $\{x_t^s\}$ such that[4]

$$\begin{bmatrix} \Delta X_t \\ x_t^s \end{bmatrix} = \begin{bmatrix} \Lambda \\ \Lambda^s \end{bmatrix} \Delta F_t + \begin{bmatrix} \Delta e_t \\ e_t^s \end{bmatrix} = \begin{bmatrix} \Lambda \\ \Lambda^s \end{bmatrix} u_t + \begin{bmatrix} \Delta e_t \\ e_t^s \end{bmatrix}, \; (t = 2, \cdots T), \tag{8}$$

where x_t^s denotes an N_0-dimensional vector containing stationary variables, $\Lambda^s = \left[\lambda_{N_1+1}^s, \cdots, \lambda_{N_1+N_0}^s\right]'$ is the corresponding factor loadings, and $e_t^s = \left[e_{N_1+1,t}^s, \cdots, e_{N_1+N_0,t}^s\right]'$ is an $N_0 \times 1$ vector of error components for x_t^s.[5] One can estimate the space spanned by u_t using the principal component estimation method as in Bai (2003). This estimator is denoted as $\tilde{u}_t$. An advantage of employing model (8) is that additional stationary variables x_t^s can be used for forecasting.

Under Assumptions A - C in Bai (2003),

$$\min\{N, T\} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{u}_t - H_D' u_t \right\|^2 \right) = O_p(1), \tag{9}$$

where H_D is an $r \times r$ rotation matrix that corresponds to matrix H in Theorem 1 of Bai (2003). If $\frac{\sqrt{N}}{T} \to 0$ and Assumptions A - G in Bai (2003) are assumed, for each t,

$$\sqrt{N} \left(\tilde{u}_t - H_D' u_t \right) \xrightarrow{d} N(0, \left(V^D\right)^{-1} Q^D \Xi_t Q^{D\prime} \left(V^D\right)^{-1}), \tag{10}$$

where V^D and Q^D are $r \times r$ matrices[6], and

$$\Xi_t = \lim_{N\to\infty} \frac{1}{N} \left[\sum_{i=1}^{N_1} \sum_{j=1}^{N_1} E(\lambda_i \lambda_j' \Delta e_{it} \Delta e_{jt}) + \sum_{i=N_1+1}^{N} \sum_{j=N_1+1}^{N} E(\lambda_i^s \lambda_j^{s\prime} e_{it}^s e_{jt}^s) + 2 \sum_{i=1}^{N_1} \sum_{j=N_1+1}^{N} E(\lambda_i \lambda_j^{s\prime} \Delta e_{it} e_{jt}^s) \right],$$

[4] It would be more general to assume

$$\begin{bmatrix} \Delta X_t \\ x_t^s \end{bmatrix} = \begin{bmatrix} \Lambda & 0 \\ \Lambda_F^s & \Lambda_G^s \end{bmatrix} \begin{bmatrix} \Delta F_t \\ G_t \end{bmatrix} + \begin{bmatrix} \Delta e_t \\ e_t^s \end{bmatrix}, \; (t = 2, \cdots, T),$$

where x_t^s is an N_0-dimensional vector containing stationary variables, $\begin{bmatrix} \Lambda_F^s & \Lambda_G^s \end{bmatrix}$ is the corresponding factor loadings, G_t is a vector of $I(0)$ factors and e_t^s is an $N_0 \times 1$ vector of disturbances for x_t^s. But then, we cannot separate out the estimate of ΔF_t from that of G_t. The former is being used for the forecasting equation of this paper. If we use two different forecasting equations, we can accommodate the specification given above. In this case, however, it becomes difficult to compare variances of the forecasting errors. Thus, we will stick to the specification (8).

[5] It is possible to generate $I(1)$ variables from $I(0)$ variables (i.e., $X_t = \sum_{i=0}^{t} x_i^s$) and use model specification (2). However, we do not take this approach since such practice might bring $I(1)$ disturbances.

[6] Note that $H_D = \left(\frac{\Lambda'\Lambda + \Lambda^{s\prime}\Lambda^s}{N}\right)\left(\frac{u'\tilde{u}}{T-1}\right)\left(V_{NT}^D\right)^{-1}$ where $V_{NT}^D = diag\left(v_{NT,1}^D, \cdots, v_{NT,r}^D\right)$, $v_{NT,1}^D \geq \cdots \geq v_{NT,r}^D > 0$, and $v_{NT,j}^D$

which is the asymptotic variance of $\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N_1}\lambda_i\Delta e_{it}+\sum_{i=N_1+1}^{N}\lambda_i^s e_{it}^s\right)$.

Now, consider the second stage of forecasting that employs $\tilde{u}_t$. The forecasting equation is written as

$$\begin{aligned} w_{t+h} &= \mu+\alpha' H_D'^{-1}\tilde{u}_t+\beta'\Delta Z_t+\Delta\epsilon_{t+h}+\alpha' H_D'^{-1}\left(H_D' u_t-\tilde{u}_t\right) \\ &= \delta^{D\prime}\tilde{P}_t+\Delta\epsilon_{t+h}+\alpha' H_D'^{-1}\left(H_D' u_t-\tilde{u}_t\right), \end{aligned} \tag{11}$$

where $\tilde{P}_t=(1,\tilde{u}_t',\Delta Z_t')'$. Using model (11), we can obtain the OLS estimator $\hat{\delta}^D$ and the forecast of w_{T+h} is $\hat{w}_{T+h|T}=\hat{\delta}^{D\prime}\tilde{P}_T$. Thus, the forecast of the target variable y_{T+h} is $\hat{y}^D_{T+h|T}=\sum_{m=1}^{h}\hat{w}_{T+m|T}+y_T$.

Note that

$$\begin{aligned} & \hat{y}^D_{T+h|T}-y_{T+h} \\ =\ & \sum_{m=1}^{h}\hat{w}_{T+m|T}+y_T-y_{T+h} \\ =\ & \sum_{m=1}^{h}\left(\hat{w}_{T+m|T}-w_{T+m}\right) \\ =\ & (\hat{\delta}^D-\delta^D)'\left(\sum_{m=1}^{h}\tilde{P}_{T+m-h}\right)+\alpha' H_D'^{-1}\left(\sum_{m=1}^{h}\left(\tilde{u}_{T+m-h}-H_D' u_{T+m-h}\right)\right)-\sum_{m=1}^{h}\Delta\epsilon_{T+m}. \end{aligned} \tag{12}$$

This shows that the forecasting error involves terms that accumulate with h. To evaluate the asymptotic variance of $\hat{y}^D_{T+h|T}-y_{T+h}$, we introduce the following assumption that is used to derive the limiting distribution of $\hat{\delta}^D$.

Assumption 2 *Let $\Delta L_t=(1,\Delta F_t',\Delta Z_t')'$. For $h=1,2,\cdots$, the following holds.*
(i) $\frac{1}{T}\sum_{t=2}^{T-h}\Delta L_t\Delta L_t'\xrightarrow{p}\Sigma_{\Delta L}>0$.
(ii) $\frac{1}{\sqrt{T}}\sum_{t=2}^{T-h}\Delta L_t\Delta\epsilon_{t+h}\xrightarrow{d}N(0,\Sigma_{\Delta\epsilon L})$,
where $\Sigma_{\Delta\epsilon L}=\lim_{T\to\infty}\frac{1}{T}E\left(\sum_{t=2}^{T}\Delta L_t\Delta\epsilon_{t+h}\right)\left(\sum_{t=2}^{T}\Delta L_t\Delta\epsilon_{t+h}\right)'$.

Parts (i) and (ii) correspond to Assumption E of Bai and Ng (2006). Part (i) is the law of large numbers for $\frac{1}{T}\sum_{t=2}^{T-h}\Delta L_t\Delta L_t'$, and part (ii) is a central limit theorem for $\frac{1}{\sqrt{T}}\sum_{t=2}^{T-h}\Delta L_t\Delta\epsilon_{t+h}$.

---

denotes the jth-eigenvalue of $\frac{\Delta X\Delta X'}{N(T-1)}$ where $\Delta X=\left[\left[\begin{array}{c}\Delta X_2\\ x_2^s\end{array}\right],\cdots,\left[\begin{array}{c}\Delta X_T\\ x_T^s\end{array}\right]\right]'$. V^D_{NT} converges to $V^D=diag\left(v_1^D,\cdots,v_r^D\right)$ in probability, where $v_1^D>\cdots>v_r^D>0$; v_j^D denotes the j^{th}-eigenvalue of $\left(c\Sigma_\Lambda+(1-c)\Sigma_{\Lambda^s}\right)^{1/2}\Sigma_u\left(c\Sigma_\Lambda+(1-c)\Sigma_{\Lambda^s}\right)^{1/2}$ and Υ^D is the corresponding eigenvector matrix such that $\Upsilon^{D\prime}\Upsilon^D=I_r$. Then, $Q^D=\left(V^D\right)^{1/2}\Upsilon^{D\prime}\left(c\Sigma_\Lambda+(1-c)\Sigma_{\Lambda^s}\right)^{-1/2}$. Note that $c\Sigma_\Lambda+(1-c)\Sigma_{\Lambda^s}$ denotes the probability limit of $\frac{\Lambda'\Lambda+\Lambda^{s\prime}\Lambda^s}{N}$.

The following lemma will be used to derive an approximate asymptotic variance of the forecasting error $\hat{y}^D_{T+h|T} - y_{T+h}$. It follows straightforwardly from Theorem 1 of Bai and Ng (2006).

Lemma 2 *Suppose that assumptions for the asymptotic result (9) and Assumption 2 hold. In addition, assume*

$$Var\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\left(\sum_{i=1}^{N_1}\Delta Z_t \Delta e_{it}\lambda_i' + \sum_{i=N_1+1}^{N}\Delta Z_t e_{it}^s \lambda_i^{s\prime}\right)\right) = O\left(1\right). \tag{13}$$

Let $\Sigma_{\Lambda D} = plim_{N\to\infty}\frac{\Lambda'\Lambda + \Lambda^{s\prime}\Lambda^s}{N} = c\Sigma_\Lambda + (1-c)\Sigma_{\Lambda^s}$ *and* $\Psi_D = 1 \oplus \left(V^D\right)^{-1/2}\Upsilon_D'\Sigma_{\Lambda D}^{1/2} \oplus I_K$ *(see footnote 6 for the definitions of* V^D *and* Υ_D*). If* $\frac{\sqrt{T}}{N} \to 0$,

$$\sqrt{T}\left(\hat{\delta}^D - \delta^D\right) \xrightarrow{d} \Psi_D'^{-1}\Sigma_{\Delta L}^{-1}N(0, \Sigma_{\Delta\epsilon L}).$$

Condition (13) imposes a restriction on dependence among $\{\Delta Z_t\}$, $\{\Delta e_{it}, e_{it}^s\}$ and $\{\lambda_i, \lambda_i^s\}$. This assumption is a counterpart of Assumption (ii) in Lemma 1.

2.3 Approximate variances of the forecasting errors

This subsection derives approximate variances of the forecasting errors from the forecasting regressions using $\tilde{F}_t$ and $\tilde{u}_t$. For these, we introduce the following assumption.

Assumption 3 *(i)* $Var(\epsilon_t) = \sigma_\epsilon^2$ *for all* t.
(ii) $E(\epsilon_t\epsilon_s) = 0$ *for* $t \neq s$.
(iii) $E\left(\epsilon_{T+h}|\left\{\epsilon_{t+h}\right\}_{t=1}^{T-h}, \{L_t\}_{t=1}^T, \{e_t\}_{t=1}^T\right) = 0$ *for any* T.
(iv) $\{F_t, Z_t, \epsilon_t\}$ *and* $\{e_{it}\}$ *are independent for all* i.

Parts (i) and (ii) of this assumption are introduced for analytic simplicity in deriving approximate variances. Parts (iii) and (iv) are employed to show that components of the forecasting errors are asymptotically uncorrelated. For example, under part (iii), ϵ_{T+h} is uncorrelated to any measurable function of $\left\{\epsilon_{t+h}\right\}_{t=1}^{T-h}, \{L_t\}_{t=1}^T$ and $\{e_t\}_{t=1}^T$, which makes ϵ_{T+h} uncorrelated to the first part of the forecasting error of $\hat{y}^L_{T+h|T}$ given in (6).

Now, we report approximate variances of the forecasting errors.

Theorem 3 *Suppose that (i) assumptions for Lemmas 1 and 2, (ii) assumptions for relations (4) and (10), and (iii) Assumption 3 hold. Let* $avar\left(\hat{y}^L_{T+h|T} - y_{T+h}\right)$ *and* $avar\left(\hat{y}^D_{T+h|T} - y_{T+h}\right)$ *be the approximate variances of the forecasting errors given* $\{L_t\}_{t=1}^T$.

(a) If $\frac{N_1}{T^3} \to 0$,

$$avar\left(\hat{y}^L_{T+h|T} - y_{T+h}\right) = \underbrace{L_T' D_T^{-1} \Sigma_L^{-1} \Sigma_{\epsilon L} \Sigma_L^{-1} D_T^{-1} L_T}_{\textit{forecasting regression part}} + \underbrace{\frac{1}{N_1} \alpha' \left(\Sigma_\Lambda^{-1} \Sigma_T \Sigma_\Lambda^{-1}\right) \alpha}_{\textit{factor estimation part}} + \sigma^2_\epsilon \tag{14}$$

(b) If $\frac{\sqrt{N}}{T} \to 0$,

$$\begin{aligned} avar\left(\hat{y}^D_{T+h|T} - y_{T+h}\right) &= \underbrace{\frac{1}{T}\left(\sum_{m=1}^h \Delta L_{T+m-h}\right)' \Sigma_{\Delta L}^{-1} \Sigma_{\Delta \epsilon L} \Sigma_{\Delta L}^{-1} \left(\sum_{m=1}^h \Delta L_{T+m-h}\right)}_{\textit{forecasting regression part}} \\ &\quad + \underbrace{\frac{1}{N} \alpha' \left(\Sigma_{\Lambda D}^{-1} \left(\sum_{m=1}^h \Xi_{T+m-h}\right) \Sigma_{\Lambda D}^{-1}\right) \alpha}_{\textit{factor estimation part}} + 2\sigma^2_\epsilon. \end{aligned} \tag{15}$$

There are three components in the forecasting error $\hat{y}^L_{T+h|T} - y_{T+h}$ as shown in equation (6). These involve $\hat{\delta}^L - \delta^L$, $\tilde{F}_T - H_L' F_T$ and ϵ_{T+h}. For $\hat{y}^D_{T+h|T} - y_{T+h}$, we have similar components. Assumption (i) in Theorem 3 is employed for the asymptotic distributions of $D_T\left(\hat{\delta}^L - \delta^L\right)$ and $\sqrt{T}\left(\hat{\delta}^D - \delta^D\right)$. Assumption (ii) in Theorem 3 is required for the asymptotic distributions of $\sqrt{N_1}\left(\tilde{F}_t - H_L' F_t\right)$ and $\sqrt{N}\left(\tilde{u}_t - H_D' u_t\right)$. The ratio restrictions, $\frac{N_1}{T^3} \to 0$ for the level-based forecasting and $\frac{\sqrt{N}}{T} \to 0$ for the differencing-based one, are needed to avoid the degenerated asymptotic distributions of the estimated factor spaces. Assumption 3 is employed to calculate the variances of the last components of the forecasting errors and to show that components of the forecasting errors are asymptotically uncorrelated.

Some interpretations of this theorem are in order. First, results (14) and (15) are related to Theorem 3 and Corollary 1 of Bai and Ng (2006). In fact, the approximate variance (15) is derived using the same method as in Bai and Ng (2006). However, the approximated variance of $\hat{y}^D_{T+h|T} - y_{T+h}$ includes terms that accumulate with h since we difference the forecasting equation to have a stationary target variable.

Second, we have $\frac{avar\left(\hat{y}^L_{T+h|T} - y_{T+h}\right)}{avar\left(\hat{y}^D_{T+h|T} - y_{T+h}\right)} \xrightarrow{p} \frac{1}{2}$ as $N, T \to \infty$, which implies that $\hat{y}^L_{T+h|T}$ is more accurate in the limit than $\hat{y}^D_{T+h|T}$, although admittedly this might be an oversimplification.

Third, we obtain a hint at what roles N_1 and N play in determining the asymptotic variances from Theorem 3. If $N_1 = N$ (i.e., all the variables are $I(1)$) and $\{e_t\}$ is a vector white noise process, $\sum_{m=1}^h \Xi_{T+m-h} -$

$\Sigma_T > 0$, which implies the second term on the right-hand-side of equation (15) is larger than that of (14). If the number of $I(1)$ variables N_1 is smaller than the total number of variables N, the magnitudes of the second terms in the approximate variances rely on the size of N_1. The larger N_1 is, the smaller the second term in (14) becomes. In other words, $\hat{y}^L_{T+h|T}$ tends to be accurate when the majority of the variables is $I(1)$.

Fourth, since it seems difficult to compare the first terms on the right-hand-sides of equations (14) and (15), we provide a simple example here that helps understanding the implications of the first terms. To this ends, assume a very simple model

$$y_{t+h} = F_t + \epsilon_{t+h}, \ (t = 1, \cdots, T-h),$$

where $\epsilon_{t+h} \backsim i.i.d. \left(0, \sigma^2_\epsilon\right)$. Note that $L_t = F_t$ for each t, $D_T = T$ and $\alpha = 1$. Thus, we have $\Sigma_L = \int_0^1 B^2_F(s)\, ds$ and $\Sigma_{\epsilon L} = \sigma^2_\epsilon \int_0^1 B^2_F(s)\, ds$, from which we deduce

$$L'_T D_T^{-1} \Sigma_L^{-1} \Sigma_{\epsilon L} \Sigma_L^{-1} D_T^{-1} L_T = \frac{\sigma^2_\epsilon F^2_T}{T^2 \int_0^1 B^2_F(s)\, ds}. \tag{16}$$

Hence, the first term of $avar\left(\hat{y}^L_{T+h|T} - y_{T+h}\right)$ is of $O_p\left(\frac{1}{T}\right)$ since $\frac{1}{T}F_T = O_p\left(\frac{1}{\sqrt{T}}\right)$. Next, we consider the first term of $avar\left(\hat{y}^D_{T+h|T} - y_{T+h}\right)$. Since $\Sigma_{\Delta\epsilon L} = 2\sigma^2_\epsilon \Sigma_{\Delta L}$ if $\{u_t\}$ and $\{\epsilon_{t+h}\}$ are independent, we have

$$\frac{1}{T}\left(\sum_{m=1}^h \Delta L_{T+m-h}\right)' \Sigma_{\Delta L}^{-1} \Sigma_{\Delta\epsilon L} \Sigma_{\Delta L}^{-1} \left(\sum_{m=1}^h \Delta L_{T+m-h}\right) = \frac{2\sigma^2_\epsilon \left(\sum_{m=1}^h u_{T+m-h}\right)^2}{T\Sigma_{\Delta L}} = O_p\left(\frac{1}{T}\right). \tag{17}$$

Hence, (16) and (17) have the same asymptotic magnitude. However, $\left(\sum_{m=1}^h u_{T+m-h}\right)^2$ may increase with h, which suggests that the first term of $avar\left(\hat{y}^D_{T+h|T} - y_{T+h}\right)$ may also do so.

Fifth, a forecasting horizon h affects the forecasting error variance of $\hat{y}^D_{T+h|T}$, but not that of $\hat{y}^L_{T+h|T}$. We may infer from this that the forecasting error variance of $\hat{y}^D_{T+h|T}$ grows with h, while that of $\hat{y}^L_{T+h|T}$ does not. This implies that the factor-based forecasts using non-differenced data may have an advantage in the long-run forecasting. In fact, this is what we would observe in the simulation results of the next section.

Sixth, if we assume that $\{\epsilon_t\}$ is a weakly stationary process, $2\sigma^2_\epsilon$ in (15) should be replaced by $Var(\epsilon_{T+h} - \epsilon_T)$. Additionally, if Assumption 3 (iii), (iv) are eliminated, approximate asymptotic covariances among the components of the forecasting errors should be added to the approximate variances in Theorem 3.

3 Simulations

This section compares the forecasting performance of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ by simulations. The data for $\{X_{it}\}$ were generated by the following data generating processes (DGPs):

$$\begin{aligned} X_{it} &= \lambda_i' F_t + \sigma_{xi} e_{it} \text{ (for } i = 1, \cdots, N_1); \\ x^s_{it} &= \lambda_i^{s\prime} u_t + \sigma_{xi} e_{it} \text{ (for } i = N_1+1, \cdots, N); \\ \lambda_i &\backsim i.i.d.\ N(0, I_r) \text{ (for } i = 1, \cdots, N_1);\ \lambda_i^s \backsim i.i.d.\ N(0, I_r) \text{ (for } i = N_1+1, \cdots, N); \\ F_t &= F_{t-1} + u_t,\ u_t \backsim i.i.d.\ N(0, I_r) \text{ and } u_0 = 0; \\ e_{it} &= \rho_i e_{i,t-1} + \sigma_{\xi i} \xi_{it},\ \xi_{it} \backsim i.i.d.\ N(0,1) \text{ for } i = 1, \cdots, N, \end{aligned} \tag{18}$$

where $\sigma^2_{xi} = r(1-\rho_i^2)/\sigma^2_{\xi i}$, $\{\rho_i\}$ was taken from $U[0.3, 0.8]$ and $\{\sigma^2_{\xi i}\}$ from $U[1,3]$. For $\{F_t\}$ and $\{e_{it}\}$, we generated $T+60$ observations for each i and employed the last T observations for each series. Since $Var\left(\lambda_i' F_t\right) = tr$ and $Var\left(\sigma_{xi} e_{it}\right) = \frac{r(1-\rho_i^2)}{\sigma^2_{\xi i}} \frac{\sigma^2_{\xi i}}{(1-\rho_i^2)} = r$, the signal-to-noise ratio is t for all the parameter values. For the sample size, we considered $(T,N) = (50, 80)$, $(50, 150)$, $(150, 80)$ and $(150, 150)$ with three cases of $\frac{N_1}{N}$ $(= \frac{1}{10}, \frac{1}{2}$, and $\frac{9}{10})$. For the number of factors, we set r to be 1, 3 and 5.

Next, the forecasting equation in levels was generated by using $\{F_t\}$ from DGP (18):

$$y_{t+h} = \mu + \alpha' F_t + \beta y_t + \epsilon_{t+h},\ (t = 1, \cdots, T-h). \tag{19}$$

Throughout the simulation, we set $(\mu, \alpha', \beta) = \left(0.5, \underbrace{1, \cdots, 1}_{r \text{ times}}, 0.2\right)$.[7] For forecasting errors, we considered the $AR(1)$ process: $\epsilon_t = \rho_\epsilon \epsilon_{t-1} + \omega_t$ where $\omega_t \backsim i.i.d.\ N(0,1)$. We set ρ_ϵ to be 0 and 0.5. The second case represents weakly correlated forecasting errors. For a forecasting horizon h, we considered $h = 1$, 3 and 12. The target variable y_{t+h} can also be forecasted by using the model

$$\Delta y_{t+m} = \alpha' u_t + \beta \Delta y_t + \Delta \epsilon_{t+h},\ (t = 2, \cdots, T-h; m = 1, \cdots, h). \tag{20}$$

Then, $\hat{y}^D_{T+h|T} = y_T + \sum_{m=1}^h \Delta \hat{y}_{T+m|T}$ where $\Delta \hat{y}$ denotes the forecast of Δy. For this model, $\{u_t\}$ is estimated by using $\{\Delta X_{it}, x^s_{it}\}$.

Table 1 reports empirical RMSFEs of the two forecasts. The number of iterations for the calculation of the empirical mean squared forecasting errors was set to be 2,000. Let $RMSFE^L_h$ and $RMSFE^D_h$ denote

[7] In the supplement to this paper, we also reported simulation results for $(\mu, \beta) = (1, 1)$. It represents a stronger signal than $(\mu, \beta) = (0.5, 0.2)$. Overall simulation results in that case are mostly similar to those in the main draft. If a signal is stronger, however, some efficiency gains in $\hat{y}^L_{T+h|T}$ are observed. The relative efficiency of $\hat{y}^L_{T+h|T}$ tends to improve with a stronger signal.

respectively empirical RMSFE from models (1) and (7). To illustrate relative efficiency of the two forecasts, we report the ratio of $RMSFE$s defined by $\frac{L}{D} = \frac{RMSFE_h^L}{RMSFE_h^D}$. Results in Table 1 are summarized as follows.

[Table 1 here]

(a) In almost all the cases, $\hat{y}_{T+h|T}^L$ performs better than $\hat{y}_{T+h|T}^D$. When we have small $\frac{N_1}{N}$ and h (notably $\frac{N_1}{N} = 0.1$ and $h = 1$) and large r, $\hat{y}_{T+h|T}^D$ sometimes performs better than $\hat{y}_{T+h|T}^L$. These confirm the large-sample efficiency advantage of $\hat{y}_{T+h|T}^L$ reported in Theorem 3.

(b) Both $RMSFE_h^L$ and $RMSFE_h^D$ tend to decrease as N increases at the same value of T. It means that a larger number of cross-sectional units leads to better performance of forecasting at least according to our simulation design. If $\{\epsilon_{t+h}\}$ is weakly correlated and T is large, it seems that $RMSFE_h^L$ does not show a decreasing pattern when N increases.

(c) As h increases, so does $RMSFE_h^D$, while $RMSFE_h^L$ behaves stably. We observe from $\frac{L}{D}$ that the relative efficiency gain of $\hat{y}_{T+h|T}^L$ improves as h increases. This is in accordance with the contents of Theorem 3 and indicates the advantage of $\hat{y}_{T+h|T}^L$ for the long-run forecasting.

(d) Even for very small $\frac{N_1}{N}$ (i.e., $\frac{N_1}{N} = 0.1$), $\hat{y}_{T+h|T}^L$ usually performs better than $\hat{y}_{T+h|T}^D$.

(e) $RMSFE_h^L$ tends to decrease as $\frac{N_1}{N}$ increases although there are some exceptions at $(T, N) = (150, 150)$ (large N and T).

(f) Both $RMSFE_h^L$ and $RMSFE_h^D$ increase as does the true number of factors r. The relative efficiency gain of $\hat{y}_{T+h|T}^L$ (represented by $\frac{L}{D}$) tends to increase as does r, although there are exceptions especially at $h = 1$. Also, the pattern of improving efficiency gain of $\hat{y}_{T+h|T}^L$ corresponding to increasing $\frac{N_1}{N}$ becomes clearer under large r. This indicates that using $\hat{y}_{T+h|T}^L$ has an advantage in case of large r.

(g) When $\{\epsilon_{t+h}\}$ is weakly correlated, the relative efficiency gain of $\hat{y}_{T+h|T}^L$ decreases. Even for persistent $\{\epsilon_{t+h}\}$, however, our level approach always performs better at long forecasting horizons.

In the supplement to this paper, we report additional simulation results for other DGPs: (i) highly persistent forecasting errors ϵ_{t+h}, and (ii) $I(1)$ idiosyncratic components e_{it} in the factor model. When $\{\epsilon_{t+h}\}$ is highly persistent, the relative efficiency of $\hat{y}_{T+h|T}^L$ decreases. In this case, however, $\hat{y}_{T+h|T}^L$ still has an advantage for the long-run forecasting. If $\{e_{it}\}$ is $I(1)$ and T is small ($T = 50$), $\hat{y}_{T+h|T}^L$ tends to show better performance than $\hat{y}_{T+h|T}^D$. Under large T ($T = 150$), however, we observe that the conventional approach performs better than our level approach.

In a nutshell, under DGPs (18), (19) and (20), $\hat{y}_{T+h|T}^L$ tends to perform substantially better than $\hat{y}_{T+h|T}^D$.

In particular, $\hat{y}^L_{T+h|T}$ renders more accurate predictions than $\hat{y}^D_{T+h|T}$ for the long-run forecasting. If a large numbers of $I(0)$ predictors exist, or if $\{\epsilon_{t+h}\}$ is highly persistent, $\hat{y}^D_{T+h|T}$ can be preferred to $\hat{y}^L_{T+h|T}$ at a short forecasting horizon.

4 Empirical applications

This section illustrates how to apply the two forecasting methods to real data. We will comprehensively compare performance of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ for predicting $I(1)$ monthly U.S. macroeconomic variables. As a benchmark for the comparison of forecasting performance of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$, we employ empirical $RMSFE$s. Values of the ratio of $RMSFE$s, $\frac{L}{D} = \frac{RMSFE^L_h}{RMSFE^D_h}$, are presented for comparison. To see whether the difference between $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ is statistically significant, we employ the Diebold-Mariano (DM) test statistic (Diebold and Mariano, 1995) and its p-values even though the validity of the DM test has not been established for factor-augmented forecasting models. Gonçalves, McCracken, and Perron (2017) show that testing for equal accuracy is reasonable for comparisons of nested models discussed in Clark and McCracken (2015). For forecasting horizons, we consider $h = 1, 3, 6, 12$, and 24.

We will employ the FRED-MD dataset[8] containing 123 macroeconomic variables (i.e., $N = 123$) for our forecasting exercises. 134 monthly U.S. indicators are included in the FRED-MD dataset, which is compatible with Stock and Watson's (2005). Among them, we employ 123 indicators to have a balanced panel data set. For the descriptions of this data set, see McCracken and Ng (2016). The 123 monthly time series are available from January, 1960 to October, 2018. Based on the transformation rules given by McCracken and Ng (2016), we categorize the 123 variables into three groups: 35 $I(2)$ variables, 68 $I(1)$ variables, and 20 $I(0)$ variables. For the forecasting equation in differences, we use 123 transformed stationary time series. For the forecasting equation in levels, the 68 $I(1)$ variables and the 35 $I(2)$ variables are employed. The $I(2)$ variables are differenced once. That is, $\frac{N_1}{N}$ is about 0.8374.

We do forecasting exercises for three sampling periods: (i) Period I: the pre-Great Moderation period (1960-1983), (ii) Period II: the Great Moderation period (1984-June, 2007), and (iii) Period III: the crisis and aftermath period (July, 2007-October, 2018). The motivation of considering the three periods is that many empirical results advocate the existence of structural changes: from the onset of Period II, the U.S. economy became more stable; and economic uncertainty tended to increase from the beginning of Period

[8]The dataset can be found in the following website: https://research.stlouisfed.org/econ/mccracken/fred-databases/.

III but returned to the Great Moderation after 1~2 years. For each period's data environment, hence, there might exist a specific forecasting model that we need to find. By considering the three sampling periods, we hope to study how the performance of the two forecasting methods changes across the three sampling periods. The initial sampling periods for forecasting were set at 13 years for Periods I and II, and 6 years for Period III. For example, we compute the two forecasts (level-based and differencing-based) for January, 1973 using the data with the sampling period January, 1960 - December, 1972 if Period I and $h = 1$ are chosen. Then, we compute the two forecasts for February, 1973 using the data with the sampling period February, 1960 - January, 1973. This procedure is repeated for the calculations of $RMSFE$s.

When we estimate the number of factors using the full sample, Bai's (2004) IPC_1 criterion detects 6 factors. We use IPC_1 since it performs best in his simulation study. However, there is no theoretical foundation whether this criterion is effective for the factor-augmented forecasting equation. Therefore, by performing cross-validation, we recursively determine the effective autoregressive lags (p_f^*) as well as the numbers of factors (r_f^*) in the forecasting equation. Using presamples, we consider various forecasting models across the numbers of factors (r_f) and autoregressive lags (p_f) and choose a model yielding the best prediction.[9]

We checked stationarity of the estimated idiosyncratic components by conducting unit root tests. This step is required to verify whether the real data set satisfies our theoretical assumptions. We observe that the KPSS test (introduced by Kwiatkowski, Phillips, Schmidt, and Shin, 1992) does not reject stationarity for all the variables while the augmented Dickey Fuller (ADF) test does not reject the unit root hypothesis for 49 variables (among 103 variables). In the supplement to this paper, we report the testing results in Table 4 and their interpretations in the first subsection of Appendix III.

In the following subsections, we report the results of our forecasting exercises. First, we consider forecasting 68 $I(1)$ monthly U.S. macroeconomic variables and show overall relative performance (efficiency) of

[9] To generate each forecast $\hat{y}^L_{T+h|T}$, first, we withhold half of time series observations ($\{y_s\}^T_{s=[\frac{T}{2}]+1}$) from the model estimation. Then, for each (r_f, p_f), we recursively generate the forecasts $\left\{\hat{y}^L_{[\frac{T}{2}]+h|[\frac{T}{2}]}, \cdots, \hat{y}^L_{T|T-h}\right\}$ and compute ex-post forecasting errors $\left\{y_{[\frac{T}{2}]+h} - \hat{y}^L_{[\frac{T}{2}]+h|[\frac{T}{2}]}, \cdots, y_T - \hat{y}^L_{T|T-h}\right\}$. By using these, for each (r_f, p_f), we evaluate the root mean squared ex-post forecasting error. The combinations of $r_f = 0, \cdots, 6$ and $p_f = 0, \cdots, 20$ are considered in the forecasting exercises. That is, the total of 147 forecasts are compared in order to generate a forecast. For the case of $h = 24$ and Period 3, we set the maximum p_f to be 11 since the time series observations for Period 3 are smaller than those for other periods. We choose an optimal pair (r_f^*, p_f^*), which yields the best prediction in terms of the root mean squared ex-post forecasting errors. The same strategy is used for $\hat{y}^D_{T+h|T}$.

the level-based forecasts. Next, we choose two representative variables, the CPI-based U.S. inflation rate and the Industrial Production Index (hereafter, IP), and analyze their forecasts in detail.

4.1 Overall forecasting results for 68 $I(1)$ U.S. macroeconomic variables

As Diebold and Kilian (2000) point out, many monthly macroeconomic variables are highly persistent. Thus, our level-based forecasting method seems to be adequate for those variables. In this subsection, we report the results of forecasting exercises for 68 $I(1)$ monthly variables. The $I(1)$ variables consist of six groups of series as classified by McCracken and Ng (2016): (i) Output and income, (ii) Labor market, (iii) Consumption, orders, and inventories, and (iv) Money and credit, (v) Interest and exchange rates, and (vi) Stock market. We consider past target variables, and time trend[10] as regressors $\{Z_t\}$. A constant term μ is included in every forecasting equation throughout this section.

Following Marcellino, Stock, and Watson (2006), we report summary statistics of the empirical distributions of $\frac{L}{D}$ in Table 2 and plots of them in Figure 1. For these, we computed $\frac{L}{D}$ for each $I(1)$ series and obtained 68 $\frac{L}{D}$ values at each forecasting horizon. Using these, we computed the mean, minimum, maximum and percentiles of $\frac{L}{D}$. To investigate whether the two forecasts are significantly different or not, we report the relative frequencies of the events $\{DM > 1.645\}$ and $\{DM < -1.645\}$. If $DM > 1.645$, the null hypothesis that there is no difference between the two forecasts is rejected at the 5% level against the alternative that $\hat{y}^L_{T+h|T}$ performs better than $\hat{y}^D_{T+h|T}$. On the other hand, we decide that $\hat{y}^D_{T+h|T}$ outperforms $\hat{y}^L_{T+h|T}$ if $DM < -1.645$. We denote the relative frequencies as $\widehat{\Pr}\{DM > 1.645\}$ and $\widehat{\Pr}\{DM < -1.645\}$.

[Table 2 here]

[Figure 1 here]

The results reported in Table 2 and Figure 1 are summarized as follows.

(a) In most entries, we observe that the basic statistics and the empirical percentiles for $\frac{L}{D}$ take values smaller than 1, which means that $\hat{y}^L_{T+h|T}$ performs better than $\hat{y}^D_{T+h|T}$.

(b) The rejection frequency rates of the DM test also suggest that $\hat{y}^L_{T+h|T}$ tends to forecast better than $\hat{y}^D_{T+h|T}$.

[10] We find the presence of a time trend in most of the $I(1)$ variables. Hence, we include a time trend in the level-based forecasting equation. For the case of no time trend in an $I(1)$ variable, see the forecasting exercise of the U.S. inflation rate in the next subsection.

(c) Overall, the relative efficiency gain of $\hat{y}^L_{T+h|T}$ is largest in Period III while it is smallest in Period I. The rejection rates of the DM test also show that $\hat{y}^L_{T+h|T}$ performs best in Period III and worst in Period I. These results imply that using the level-data for forecasting is most apposite to Period III (the crisis and aftermath period).

(d) All the basic statistics except maximum and all the empirical percentiles indicate that the efficiency gain of $\hat{y}^L_{T+h|T}$ tends to increase as does a forecasting horizon h. This means that using the level data is particularly advantageous for the long-run forecasting. However, as indicated by the maximum values of $\frac{L}{D}$, there are some examples that contradict this statement. Some of these are (i) IP: Business Equipment in Period I, (ii) Unfilled Orders for Durable Goods in Period I, (iii) Civilian Unemployed - 15 Weeks & Over in Period II, and (iv) Civilian Labor Force in Period III.

The empirical facts that $\hat{y}^L_{T+h|T}$ tends to perform better than $\hat{y}^D_{T+h|T}$ are in accordance with our theoretical finding reported in Theorem 3 and the simulation results of the previous section. In the next subsection, we report the results of forecasting exercises for two representative time series: (i) the CPI-based U.S. inflation rate, and (ii) the U.S. IP index.

4.2 Forecasting results for the U.S. inflation rate and IP index

In forecasting U.S. and U.K. inflation rates, the targeted variable y_t turns out to be $I(1)$ as shown in Stock and Watson (1999), Canova (2007) and McCracken and Ng (2016).[11] According to Stock and Watson (2005) and McCracken and Ng (2016), the logarithm of CPI is treated as an $I(2)$ variable. King and Watson (2012) document that the U.S. inflation rate has become less persistent since the beginning of the Great Moderation. We also verify this in our dataset. For the U.S. inflation rate, its standard deviation is largest in Period I, smallest in Period II, and increases again in Period III.

For regressors $\{Z_t\}$ in forecasting the U.S. inflation rate, three types of variables are considered: (i) past U.S. inflation rates, (ii) the unemployment rate, and (iii) the term spread.[12] We do not include a linear time trend in the model since we find that there is no time trend in the U.S. inflation rate. The unemployment rate is often used to forecast inflation rates because of their stable relationship predicted by the Phillips curve. Although some articles reject the stable relationship empirically (cf. Atkeson and Ohanian, 2001;

[11] There are various works in predicting inflation rates using estimated factors from a large macroeconomic dataset. Examples are Stock and Watson (2002, 2003) for U.S. inflation; Artis et al. (2005) for U.K. inflation; Shintani (2005) for Japanese inflation; Hofmann (2009) for euro-area inflation.

[12] For $\hat{y}^D_{T+h|T}$, regressors were differenced to be stationary.

Fisher, Liu and Zhou, 2002; Ang, Bekaert and Wei, 2007; Stock and Watson, 2008; Dotsey, Fujita and Stark, 2015), we still employ the unemployment rate as an explanatory variable considering the importance of the Phillips curve.[13] The expectations hypothesis of interest rates' term structure indicates that the term spread contains some information regarding future inflation rates. Mishkin (1990) reports some empirical evidence that changes in inflation rates can be forecasted by the term spread. Although Stock and Watson (2003) argue that interest rates' term structure fails to deliver convincing evidence for predictability of inflation rates when lagged inflation rates are present as regressors (e.g., Kozicki, 1997; Estella and Mishkin, 1997), we will consider the term spread as an additional explanatory variable. The term spread is calculated as the difference between the yields on 10-year and 3-month Treasury securities.

The first part of Table 3 reports results of the out-of-sample forecasting exercises for the U.S. inflation rate.

[Table 3 here]

The inflation forecasting results in Table 3 are summarized as follows.

(a) For all the sampling periods, $\hat{y}^L_{T+h|T}$'s performance does not vary much along with h compared to that of $\hat{y}^D_{T+h|T}$. When h increases, the relative efficiency of $\hat{y}^L_{T+h|T}$ improves. $RMSFE^D_h$ tends to increase with h. This feature is more pronounced in Periods I and III than in Period II.

(b) Except at $h = 1$ in Period II (Great Moderation), $\hat{y}^L_{T+h|T}$ outperforms $\hat{y}^D_{T+h|T}$.

(c) We observe that $RMSFE$s of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ show heterogeneous features across the three sampling periods. In Period I, $RMSFE$s of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ tend to be higher than those in Periods II and III (except at $h = 24$). It might reflect the volatile nature of the U.S. economy before the Great Moderation. In Period I, we find that using $\hat{y}^L_{T+h|T}$ is more beneficial than using $\hat{y}^D_{T+h|T}$ since $\hat{y}^L_{T+h|T}$ shows robustness to different forecasting horizons.

(d) For the Great Moderation period (Period II), $RMSFE$s of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ decrease relative to those in Period I. During the Great Moderation period, $RMSFE^D_h$ shows relatively stable behavior compared to those in Periods I and III. At $h = 1$ for Period II, using $\hat{y}^D_{T+h|T}$ has an advantage in terms of relative efficiency. It seems that price variables became less volatile since the onset of the Great Moderation.

(e) For the crisis and aftermath period (Period III), $\hat{y}^L_{T+h|T}$ performs well in terms of $RMSFE$ and sensitivity to h. Among the three sampling periods, $RMSFE^L_h$s in Period III are smallest. Forecasting

[13] Stock and Watson (1999) use other variables that signal real economic activities as alternatives to the unemployment rate, but they are not used here because they do not bring significantly different results.

performance of $\hat{y}^D_{T+h|T}$ becomes worse when h increases.

(f) At the 5% significance level, the DM test rejects the null hypothesis of equivalent performance of the two forecasts (for both one-sided and two-sided tests) except for the case of $h = 1$, and that of $h = 24$ in Period II. These results imply that the better performance of $\hat{y}^L_{T+h|T}$ relative to $\hat{y}^D_{T+h|T}$ is statistically significant in most cases.

(g) To study statistical properties of forecasting errors, we conducted a unit root test and estimated $AR(1)$ coefficients using the residuals in equation (1), the results of which are reported in Table 4. For all the sampling periods and all h, we reject the null hypothesis of a unit root at the 5% significance level. The maximum value of the estimated $AR(1)$ coefficients is 0.4440. In the Great Moderation period (Period II), the values of the estimated $AR(1)$ coefficients are lower than those in other periods. We also find that the values of the estimated $AR(1)$ coefficients tend to increase as h increases.

In forecasting the U.S. inflation, we observe that $\hat{y}^L_{T+h|T}$ performs better than $\hat{y}^D_{T+h|T}$ except at $h = 1$. This result is in accordance with the conclusion of our simulation study which states that $\hat{y}^L_{T+h|T}$ tends to perform substantially better than $\hat{y}^D_{T+h|T}$ and that $\hat{y}^D_{T+h|T}$ has an advantage relative to $\hat{y}^L_{T+h|T}$ at $h = 1$ when both the numbers of factors and $I(0)$ variables are large. The efficiency gain of our method for the U.S. inflation rate is largest in the crisis and aftermath period while it is smallest in the Great Moderation period.

In the forecasting literature, the U.S. IP index is differenced once to apply the conventional differencing-based forecasting model (e.g., McCracken and Ng, 2016). According to Stock and Watson (2005) and McCracken and Ng (2016), the logarithm of the IP index is treated as an $I(1)$ variable. For the U.S. IP, we observe a decreasing pattern in its standard deviation across all the time periods. Using the same framework as for the U.S. inflation, we conducted forecasting exercises for the U.S. IP index. The target variable in this exercise is the logarithm of the U.S. IP index. Since it seems that the target variable has an upward trend, we include a linear time trend in $\{Z_t\}$ in addition to past dependent variables and estimated factors.

The IP index forecasting results (see the second part of Table 3) are summarized as follows.

(a) Except at $h = 3, 6$ in Period I, $\hat{y}^L_{T+h|T}$ performs better than $\hat{y}^D_{T+h|T}$. In Periods II and III, it seems that the relative efficiency of $\hat{y}^L_{T+h|T}$ tends to increase with h. However, this feature is not obvious in Period I.

(b) In the pre-Great Moderation period (Period I), $RMSFE$s of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ are higher than those in Periods II and III. In the Great Moderation period (Period II), $RMSFE$s of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ decrease relative to those in Period I. For the crisis and aftermath period (Period III), we observe that $RMSFE$s of $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ are smallest among the three sampling periods.

(c) We reject the null hypothesis of the DM test when $h = 1, 3, 6$, and 12 in Period III and $h = 6$ in Period I. In case of $h = 6$ with Period I, the DM test statistic takes a negative value. For the other rejection cases, the DM statistics are positive. It implies that $\hat{y}^D_{T+h|T}$ significantly outperforms $\hat{y}^L_{T+h|T}$ only for the case of $h = 6$ in Period I.

(d) For the estimated residuals in the level-based forecasting equation, we reject the null hypothesis of a unit root at the 5% significance level at $h = 1$. At $h = 3, 6, 12$, and 24, we do not reject the null hypothesis (except for $h = 3$ in Period I).[14] The maximum value of the estimated $AR(1)$ coefficients is 0.9194 (at $h = 12$ in Period II). Compared to the residuals of the U.S. inflation rate, the value of the estimated $AR(1)$ coefficients for the IP index tends to be higher, indicating that they are more persistent. Like the case of the U.S. inflation rate, however, the values of the estimated $AR(1)$ coefficients tend to increase with h.

For the U.S. Industrial Production, we find that $\hat{y}^L_{T+h|T}$ usually performs better than $\hat{y}^D_{T+h|T}$ and that the relative efficiency of $\hat{y}^L_{T+h|T}$ is quite pronounced in the crisis and aftermath period and less so in the pre-Great Moderation period.

In the supplement to this paper, we report the detailed results of ex-post forecasting for each representative economic variable of each group (categorized by McCracken and Ng (2016)): (i) All employees: Total nonfarm (EMP), (ii) Real personal consumption expenditures (Consumption), (iii) Real M2 money stock (M2_Real), (iv) Effective federal fund rate (FEDFUNDS), and (v) S&P's common stock price index: Composite (S&P 500). For those variables, we also observe that $\hat{y}^L_{T+h|T}$ outperforms $\hat{y}^D_{T+h|T}$ in most cases.

5 Conclusion and further remarks

We have investigated performance of factor-based forecasts using differenced and non-differenced data. Approximate variances of the forecasting errors from the two forecasts are derived and compared. The derived approximate variances of forecasting errors reveal that the forecasts from the regression in levels tend to be more accurate than those from the regression in differences. As contrary to the conventional method based on differenced data, the ordinary least square estimator for the forecasting equation using

[14] Since unit root tests tend to have low power, this does not necessarily mean that a unit root is present in the data. See Kim and Choi (2017) for related discussions and an alternative testing procedure that employs optimal significance level.

level data is free of asymptotic bias without a ratio restriction on the number of variables and the number of time series observations for factor estimation. We have also conducted simulations to compare root mean squared forecasting errors of the two competing forecasts. Simulation results indicate that the forecasting using non-differenced data mostly performs better in terms of root mean squared forecasting errors. The advantage of using non-differenced data is more pronounced when a forecasting horizon is long and the number of factors is large. Using differenced data can be preferred if a researcher can employ many $I(0)$ variables for estimating factor spaces and a forecasting horizon is short, or if forecasting errors are highly persistent. Hence, a practical implication of this paper is that using non-differenced data is advisable for the factor-based forecasting especially for the long-run forecasting. Lastly, we have applied the two competing forecasting methods to 68 $I(1)$ monthly U.S. macroeconomic variables across a range of forecasting horizons and sampling periods. We find that forecasts using non-differenced data tend to outperform those using differenced data. For all the cases of forecasting exercises, we ascertain that the level-based forecasts show better performance in the long-run forecasting than those using differenced data.

Appendix I: Proofs of main results

Proof of Lemma 1: Using equation (5), we have

$$\begin{aligned} D_T\left(\hat{\delta}^L - \delta^L\right) &= \left(D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\tilde{L}_t' D_T^{-1}\right)^{-1} D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\epsilon_{t+h} \\ &\quad + \left(D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\tilde{L}_t' D_T^{-1}\right)^{-1} D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\left(H_L'F_t - \tilde{F}_t\right)' H_L^{-1}\alpha \\ &= \left(D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\tilde{L}_t' D_T^{-1}\right)^{-1} D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\epsilon_{t+h} + o_p(1). \end{aligned}$$

The second equality holds due to Lemma A.II.1. Moreover, it is straightforward to show as in Lemma B.7 of Choi (2017):

$$\begin{aligned} D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\tilde{L}_t' D_T^{-1} &= D_T^{-1}\left(1\oplus H_L'\oplus I_K\right)\sum_{t=1}^{T-h}L_tL_t'\left(1\oplus H_L\oplus I_K\right)D_T^{-1} + o_p(1) \\ &\xrightarrow{d} \Psi_L\Sigma_L\Psi_L', \end{aligned} \tag{A.I.1}$$

and

$$D_T^{-1}\sum_{t=1}^{T-h}\tilde{L}_t\epsilon_{t+h} = D_T^{-1}\left(1\oplus H_L'\oplus I_K\right)\sum_{t=1}^{T-h}L_t\epsilon_{t+h} + o_p(1) \xrightarrow{d} \Psi_L\Sigma_{\epsilon L}^{1/2}N(0,I), \tag{A.I.2}$$

where $\Psi_L = 1 \oplus \left(V^L\right)^{-1/2} \Upsilon_L' \Sigma_\Lambda^{1/2} \oplus I_K$ (see Proposition 3 in Bai (2004)). Because relations (A.I.1) and (A.I.2) hold jointly, applying the continuous mapping theorem yields the stated result. The independence of $\Psi_L'^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}^{1/2}$ and $N(0,I)$ follows from the given assumption. ■

Proof of Lemma 2: This follows from Bai and Ng's (2006) Theorem 1. The details are not worth reporting and omitted. ■

Proof of Theorem 3: (i) From equation (6), we obtain

$$\begin{aligned}
\hat{y}^L_{T+h|T} - y_{T+h} &= \left(\hat{\delta}^L - \delta^L\right)' \tilde{L}_T + \alpha' H_L'^{-1}\left(\tilde{F}_T - H_L' F_T\right) - \epsilon_{T+h} \\
&= \left(\left(1 \oplus H_L' \oplus I_K\right) D_T \left(\hat{\delta}^L - \delta^L\right)\right)' \left(D_T^{-1}\left(1 \oplus H_L'^{-1} \oplus I_K\right)\tilde{L}_T\right) \\
&\quad + \alpha' H_L'^{-1}\left(\tilde{F}_T - H_L' F_T\right) - \epsilon_{T+h} \\
&= A + B - \epsilon_{T+h}, \text{ say.}
\end{aligned}$$

By applying Lemma 1, we have $\left(1 \oplus H_L' \oplus I_K\right) D_T \left(\hat{\delta}^L - \delta^L\right) \xrightarrow{d} \Sigma_L^{-1}\Sigma_{\epsilon L}^{1/2} N(0,I)$. Moreover, according to (4), $\tilde{L}_T = \begin{bmatrix} 1 \\ H_L' F_T + o_p(1) \\ Z_T \end{bmatrix}$, which gives

$$\begin{aligned}
\left(1 \oplus H_L'^{-1} \oplus I_K\right) \tilde{L}_T &= \begin{bmatrix} 1 \\ F_T + o_p(1) \\ Z_T \end{bmatrix} \\
&= L_T + o_p(1).
\end{aligned}$$

Thus, given $\{L_t\}_{t=1}^T$, $avar(A) = L_T' D_T^{-1} \Sigma_L^{-1} \Sigma_{\epsilon L} \Sigma_L^{-1} D_T^{-1} L_T$. Next, consider B. Since $H_L'^{-1}\left(\sqrt{N_1}\left(\tilde{F}_T - H_L' F_T\right)\right) \xrightarrow{d} N\left(0, \Sigma_\Lambda^{-1}\Sigma_T \Sigma_\Lambda^{-1}\right)$ (cf. relation (4)), $avar(B) = \frac{1}{N}\alpha'\Sigma_\Lambda^{-1}\Sigma_T\Sigma_\Lambda^{-1}\alpha$. Note that the asymptotic distribution of A is driven by $\{L_t, \epsilon_{t+h}\}_{t=1}^{T-h}$ while that of B rests on $\{e_t\}_{t=1}^T$. Thus, by Assumption 3 (iv), A and B are asymptotically uncorrelated. Moreover, A and ϵ_{T+h} are asymptotically uncorrelated because ϵ_{T+h} is uncorrelated to any measurable function of $\{\epsilon_{t+h}\}_{t=1}^{T-h}, \{L_t\}_{t=1}^T$ and $\{e_t\}_{t=1}^T$ due to Assumption 3 (iii). B and ϵ_{T+h} are asymptotically uncorrelated due to Assumption 3 (iv). Thus, the stated result follows.

(ii) The forecast error $\hat{y}^D_{T+h|T} - y_{T+h}$ contains the term $\sum_{m=1}^h \Delta\epsilon_{T+m} = \epsilon_{T+h} - \epsilon_T$. Under Assumption 3 (iii), ϵ_{T+h} is uncorrelated to the first term of the forecast error that involves $\hat{\delta}^D - \delta^D$, the asymptotic distribution of which is driven by $\{\epsilon_t\}_{t=1}^T$. By contrast, ϵ_T is not. But this does not cause a problem because we can show that the effect of ϵ_T on the covariance between the first and last terms is asymptotically negligible by using the assumption that ϵ_T is uncorrelated to $\epsilon_1, ..., \epsilon_{T-1}$. The rest of the proof is similar to that of part (i) and omitted. ■

Appendix II: Auxiliary Lemmas

Lemma A.II.1 *Suppose that the same assumptions for Lemma 1 hold. Then,*

$$D_T^{-1} \sum_{t=1}^{T-h} \tilde{L}_t \left(H_L' F_t - \tilde{F}_t \right)' H_L^{-1} \alpha = o_p(1).$$

Proof: (i) Let $M_t = (1, F_t' H_L, Z_t')'$ and $D_T = \sqrt{T} \oplus diag(T, ..., T) \oplus E_T$, where E_T is a diagonal matrix of dimension K making $Var\left(E_T^{-1} \sum_{t=1}^T Z_t\right) = O(1)$. Using $\tilde{L}_t = \left(\tilde{L}_t - M_t\right) + M_t = \left(0, \tilde{F}_t' - F_t' H_L, 0\right)' + M_t$, write

$$\begin{aligned} & D_T^{-1} \sum_{t=1}^{T-h} \tilde{L}_t \left(H_L' F_t - \tilde{F}_t \right)' H_L^{-1} \alpha \\ = & D_T^{-1} \sum_{t=1}^{T-h} \left(\tilde{L}_t - M_t \right) \left(H_L' F_t - \tilde{F}_t \right)' H_L^{-1} \alpha - D_T^{-1} \sum_{t=1}^{T-h} M_t \left(\tilde{F}_t - H_L' F_t \right)' H_L^{-1} \alpha \\ = & A_1 - A_2, \text{ say.} \end{aligned}$$

Relation (3) implies that

$$A_1 = \begin{pmatrix} 0 \\ -\frac{1}{T} \sum_{t=1}^{T-h} \left(\tilde{F}_t - H_L' F_t \right) \left(\tilde{F}_t - H_L' F_t \right)' H_L^{-1} \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ o_p(1) \\ 0 \end{pmatrix}, \tag{A.II.1}$$

since

$$\begin{aligned} \left\| \frac{1}{T} \sum_{t=1}^{T-h} \left(\tilde{F}_t - H_L' F_t \right) \left(\tilde{F}_t - H_L' F_t \right)' H_L^{-1} \alpha \right\| & \leq \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{F}_t - H_L' F_t \right\|^2 \right) \left\| H_L^{-1} \right\| \|\alpha\| \\ & = O_p \left(\frac{1}{\min\{N_1, T^2\}} \right) \cdot O_p(1) \cdot O_p(1) = o_p(1). \end{aligned}$$

Let $A_2 = \begin{pmatrix} A_{21} \\ A_{22} \\ A_{23} \end{pmatrix} \begin{matrix} 1 \\ r \\ K \end{matrix}$. Using the identity relation (cf. (B.1) in Bai, 2004)

$$\tilde{F}_t - H_L' F_t = \left(V_{N_1 T}^L\right)^{-1} \left[\frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \gamma_{N_1}(s,t) + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} \right],$$

where $\gamma_{N_1}(s,t) = \frac{1}{N_1} E\left(e_s' e_t\right)$, $\zeta_{st} = \frac{1}{N_1} e_s' e_t - \gamma_{N_1}(s,t)$, $\eta_{st} = \frac{1}{N_1} F_s' \Lambda' e_t$, and $\xi_{st} = \frac{1}{N_1} F_t' \Lambda' e_s$, we have

$$A_{21} = \left(V_{N_1 T}^L\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \left[\frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \gamma_{N_1}(s,t) + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} \right]' H_L^{-1} \alpha$$

By Lemma A.II.2 (i-a), (ii-a), (iii-a) and (iv-a), we obtain $A_{21} = o_p(1)$. Similarly,

$$A_{22} = \left(V_{N_1 T}^L\right)^{-1} \frac{1}{T} \sum_{t=1}^{T-h} H_L' F_t \left[\frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \gamma_{N_1}(s,t) + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} \right]' H_L^{-1} \alpha$$

and it follows from Lemma A.II.2 (i-b), (ii-b), (iii-b) and (iv-b) that $A_{22} = o_p(1)$. Last, since

$$A_{23} = \left(V_{N_1 T}^L\right)^{-1} E_T^{-1} \sum_{t=1}^{T-h} Z_t \left[\frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \gamma_{N_1}(s,t) + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} + \frac{1}{T^2} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} \right]' H_L^{-1} \alpha,$$

we obtain $A_{23} = o_p(1)$ from Lemma A.II.2 (i-c), (ii-c), (iii-c) and (iv-c). This completes the proof. ■

Lemma A.II.2 *Suppose that the same assumptions for Lemma 1 hold. Then,*

(i-a) $\frac{1}{T^{5/2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s \gamma_{N_1}(s,t) = O_p\left(\frac{1}{T}\right)$.

(i-b) $\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s F_t' \gamma_{N_1}(s,t) = O_p\left(\frac{1}{T}\right)$.

(i-c) $\frac{1}{T^2} E_T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s Z_t' \gamma_{N_1}(s,t) = O_p\left(\frac{1}{T}\right)$.

(ii-a) $\frac{1}{T^{5/2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

(ii-b) $\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s F_t' \zeta_{st} = O_p\left(\frac{1}{T}\right)$.

(ii-c) $\frac{1}{T^2} E_T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s Z_t' \zeta_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

(iii-a) $\frac{1}{T^{5/2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

(iii-b) $\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s F_t' \eta_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

(iii-c) $\frac{1}{T^2} E_T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s Z_t' \eta_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

(iv-a) $\frac{1}{T^{5/2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

(iv-b) $\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s F_t' \xi_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

(iv-c) $\frac{1}{T^2} E_T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_s Z_t' \xi_{st} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$.

Proof: The proofs of these results can be found in the first section of the supplement to this paper. ■

References

Ang, A., G. Bekaert, and M. Wei (2007). Do macro variables, asset markets, or survey forecast inflation better? *Journal of Monetary Economics*, 54, 1163–1212.

Artis, M.J, A. Banerjee, and M. Marcellino (2005). Factor forecasts for the UK. *Journal of Forecasting*, 24, 279-298.

Atkeson A, Ohanian L.E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 2-11.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71, 135-172.

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

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

Bai, J. and S. Ng (2008). Large dimensional factor models. *Foundations and Trends in Econometrics*, 3, 89-163.

Banerjee, A., M. Marcellino, and I. Masten (2014). Forecasting with factor-augmented error correction models. *International Journal of Forecasting*, 30, 589-612.

Breitung, J. and I. Choi (2013). Factor Models. Handbook of Research Methods and Applications in Empirical Macroeconomics (Hashimzade, N., and Thornton, M. A. eds.), 249-265, Edward Elgar.

Breitung, J. and S. Eickmeier (2006). Dynamic factor models. *Allegemeines Statistisches Archiv.*, 90, 27-42.

Canova F. (2007). G-7 inflation forecasts: Random walk, Phillips curve or what else? *Macroeconomic Dynamics*, 11, 1-30.

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

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

Choi, I., and B.C. Ahn (1995). Testing for cointegration in a system of equations. *Econometric Theory*, 11, 952-983.

Clark, T.E., and M.W. McCracken (2015). Nested forecast model comparisons: A new approach to testing equal accuracy. *Journal of Econometrics*, 186, 160-177.

Diebold, F.X, and R.S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X, and Kilian, L. (2000). Unit-root tests are useful for selecting forecasting models. *Journal of Business and Economic Statistics*, 18, 265-273.

Dotsey, M., S. Fujita, and T. Stark (2015). Do Phillips curves conditionally help to forecast inflation? Working paper 15-16, Federal Reserve Bank of Philadelphia.

Estrella, A., and FS. Mishkin (1997). The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European central bank. *European Economic Review*, 41, 1375-1401.

Fisher, J.D.M., C.T. Liu, and R. Zhou (2002). When can we forecast inflation? *Federal Reserve Bank of Chicago Economic Perspectives*, 1Q.2002, 30-42.

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

Gonçalves, S., M.W. McCracken, and B. Perron (2017). Tests of equal accuracy for nested models with estimated factors. *Journal of Econometrics*, 198, 231-252.

Hofman, B. (2009). Do monetary indicators lead euro area inflation? *Journal of International Money and Finance*, 28, 1165-1181.

Kim, J. and I. Choi (2017). Unit roots in economic and financial time series: a re-evaluation at the decision-based significance levels. *Econometrics*, 5, 41.

King, R.G., and M.W. Watson (2012). Inflation and unit labor cost. *Journal of Money, Credit and Banking*, 44, 111-149.

Kozicki S. (1997). Predicting real growth and inflation with the yield spread. *Federal Bank of Kansas City Economic Review*, 82, 39-57.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54, 159–178.

Ludvigson, S., and S. Ng (2011). A Factor Analysis of Bond Risk Premia. In: Ullah, A., Giles, D. (Eds.), Handbook of Empirical Economics and Finance. Chapman and Hall, 313-372.

Marcellino, M., J.H. Stock, and M.W. Watson (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135, 499-526.

McCracken M.W., and S. Ng (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business and Economic Statistics*, 34, 574-589.

Mishikin F.S. (1990). What does the term structure tell us about future inflation? *Journal of Monetary Economics*, 25, 77-95.

Phillips, P.C.B., and S. Ouliaris (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58, 165-193.

Shintani, M. (2005). Nonlinear forecasting analysis using diffusion indexes: An application to Japan. *Journal of Money, Credit and Banking*, 37, 517-538.

Stock, J.H. and M.W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics*, 44, 293-335.

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

Stock, J.H. and M.W. Watson (2003). Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature*, 41 (3), 778-829.

Stock, J.H. and M.W. Watson (2005). Implications of dynamic factor models for VAR analysis. manuscript, Harvard University.

Stock, J.H. and M.W. Watson (2008). Forecasting in Dynamic Factor Models subject to Structural Instability. The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry, (Castle, J. and N. Shephard eds), 173-205, Oxford University Press.

Stock, J.H. and M.W. Watson (2011). Dynamic Factor Models. Oxford Handbook of Economic Forecasting, (Clements, M.P. and D.F. Hendry eds.), 35-60, Oxford University Press.

Table 1: Root mean squared forecasting errors

Part I.1: $(T, N) = (50, 80)$ and $\epsilon_t \sim i.i.d.N(0,1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	0.9860	0.7900	0.1	1	3	1.0724	0.8410	0.1	1	5	1.3947	1.0140
0.5	1	1	0.9855	0.7680	0.5	1	3	1.0283	0.7923	0.5	1	5	1.1727	0.8095
0.9	1	1	0.9820	0.7748	0.9	1	3	0.9579	0.7169	0.9	1	5	1.0495	0.7645
0.1	3	1	0.9840	0.6971	0.1	3	3	1.0795	0.6753	0.1	3	5	1.3930	0.7515
0.5	3	1	0.9814	0.6911	0.5	3	3	1.0126	0.6481	0.5	3	5	1.1719	0.6197
0.9	3	1	0.9551	0.6798	0.9	3	3	0.9680	0.6564	0.9	3	5	1.0278	0.6177
0.1	12	1	0.9940	0.6637	0.1	12	3	1.0624	0.4532	0.1	12	5	1.3018	0.3910
0.5	12	1	0.9406	0.6380	0.5	12	3	1.0078	0.4646	0.5	12	5	1.1402	0.3820
0.9	12	1	0.9293	0.6286	0.9	12	3	0.9561	0.5285	0.9	12	5	1.0009	0.4689

Part I.2: $(T, N) = (50, 150)$ and $\epsilon_t \sim i.i.d.N(0,1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	0.9747	0.7614	0.1	1	3	1.1437	0.8737	0.1	1	5	1.7498	1.2758
0.5	1	1	1.0086	0.7698	0.5	1	3	0.9716	0.7407	0.5	1	5	1.0489	0.7905
0.9	1	1	0.9403	0.7604	0.9	1	3	0.9559	0.7324	0.9	1	5	1.0051	0.7548
0.1	3	1	0.9913	0.7151	0.1	3	3	1.1311	0.7446	0.1	3	5	1.8125	1.0077
0.5	3	1	0.9817	0.6918	0.5	3	3	0.9846	0.6390	0.5	3	5	1.0276	0.6450
0.9	3	1	0.9534	0.6914	0.9	3	3	0.9388	0.6495	0.9	3	5	0.9958	0.6418
0.1	12	1	0.9496	0.6564	0.1	12	3	1.0720	0.4455	0.1	12	5	1.7208	0.4775
0.5	12	1	0.9354	0.6296	0.5	12	3	0.9467	0.4877	0.5	12	5	0.9985	0.4083
0.9	12	1	0.9519	0.6534	0.9	12	3	0.9548	0.5413	0.9	12	5	0.9852	0.4736

Part I.3: $(T, N) = (150, 80)$ and $\epsilon_t \sim i.i.d.N(0,1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	1.0273	0.7944	0.1	1	3	1.1855	0.8683	0.1	1	5	1.5132	1.0710
0.5	1	1	0.9993	0.7762	0.5	1	3	1.0977	0.8189	0.5	1	5	1.2454	0.8504
0.9	1	1	0.9859	0.7754	0.9	1	3	1.0344	0.7670	0.9	1	5	1.0978	0.7653
0.1	3	1	1.0142	0.7311	0.1	3	3	1.1451	0.7464	0.1	3	5	1.5379	0.8505
0.5	3	1	1.0228	0.7331	0.5	3	3	1.0781	0.7043	0.5	3	5	1.2669	0.6969
0.9	3	1	0.9614	0.6877	0.9	3	3	1.0048	0.6704	0.9	3	5	1.1003	0.6685
0.1	12	1	1.0293	0.6740	0.1	12	3	1.1709	0.5316	0.1	12	5	1.6043	0.4993
0.5	12	1	0.9605	0.6580	0.5	12	3	1.1075	0.4962	0.5	12	5	1.2875	0.4262
0.9	12	1	0.9927	0.6939	0.9	12	3	1.0160	0.6136	0.9	12	5	1.1086	0.5479

Part I.4: $(T, N) = (150, 150)$ and $\epsilon_t \sim i.i.d.N(0,1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	1.0078	0.7815	0.1	1	3	1.2020	0.8850	0.1	1	5	1.8652	1.3208
0.5	1	1	0.9706	0.7610	0.5	1	3	1.0238	0.7599	0.5	1	5	1.0991	0.7970
0.9	1	1	0.9829	0.7803	0.9	1	3	1.0284	0.7726	0.9	1	5	1.0474	0.7684
0.1	3	1	0.9723	0.6934	0.1	3	3	1.2044	0.7730	0.1	3	5	1.0474	0.7684
0.5	3	1	0.9846	0.7036	0.5	3	3	1.0272	0.6733	0.5	3	5	1.1329	0.7007
0.9	3	1	1.0036	0.7175	0.9	3	3	1.0028	0.6759	0.9	3	5	1.0875	0.6850
0.1	12	1	1.0105	0.6909	0.1	12	3	1.2352	0.5812	0.1	12	5	2.1157	0.6719
0.5	12	1	0.9688	0.6864	0.5	12	3	1.0407	0.5447	0.5	12	5	1.1328	0.4686
0.9	12	1	0.9688	0.6864	0.9	12	3	0.9854	0.6055	0.9	12	5	1.0360	0.5203

Part II.1: $(T, N) = (50, 80)$ and $\epsilon_t = 0.5\epsilon_{t-1} + \omega_t$ where $\omega_t \sim i.i.d.N(0, 1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	1.0864	0.9790	0.1	1	3	1.1497	1.0338	0.1	1	5	1.3634	1.1439
0.5	1	1	1.0398	0.9242	0.5	1	3	1.1274	0.9679	0.5	1	5	1.2080	0.9482
0.9	1	1	1.0525	0.9585	0.9	1	3	1.0755	0.9573	0.9	1	5	1.0732	0.9049
0.1	3	1	1.1385	0.7679	0.1	3	3	1.1804	0.7501	0.1	3	5	1.4657	0.7830
0.5	3	1	1.1108	0.7681	0.5	3	3	1.1413	0.7072	0.5	3	5	1.2299	0.6782
0.9	3	1	1.1090	0.7434	0.9	3	3	1.1180	0.7201	0.9	3	5	1.1312	0.6650
0.1	12	1	1.1020	0.6954	0.1	12	3	1.1133	0.4805	0.1	12	5	1.4038	0.4210
0.5	12	1	1.0732	0.6748	0.5	12	3	1.1097	0.5039	0.5	12	5	1.1565	0.3882
0.9	12	1	1.0648	0.6900	0.9	12	3	1.0377	0.5862	0.9	12	5	1.0782	0.5025

Part II.2: $(T, N) = (50, 150)$ and $\epsilon_t = 0.5\epsilon_{t-1} + \omega_t$ where $\omega_t \sim i.i.d.N(0, 1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	1.0857	0.9732	0.1	1	3	1.2266	1.0922	0.1	1	5	1.7372	1.5298
0.5	1	1	1.0630	0.9709	0.5	1	3	1.0937	0.9935	0.5	1	5	1.1237	0.9579
0.9	1	1	1.0597	0.9697	0.9	1	3	1.0636	0.9248	0.9	1	5	1.0696	0.9366
0.1	3	1	1.1054	0.7760	0.1	3	3	1.2534	0.7692	0.1	3	5	1.9264	1.0633
0.5	3	1	1.0848	0.7582	0.5	3	3	1.0900	0.7069	0.5	3	5	1.1153	0.6855
0.9	3	1	1.0902	0.7528	0.9	3	3	1.0624	0.7110	0.9	3	5	1.0667	0.6663
0.1	12	1	1.1017	0.6948	0.1	12	3	1.1836	0.4992	0.1	12	5	1.7661	0.5165
0.5	12	1	1.0856	0.6868	0.5	12	3	1.0632	0.5379	0.5	12	5	1.0536	0.4409
0.9	12	1	1.0621	0.6918	0.9	12	3	1.0361	0.5960	0.9	12	5	1.0113	0.5105

Part II.3: $(T, N) = (150, 80)$ and $\epsilon_t = 0.5\epsilon_{t-1} + \omega_t$ where $\omega_t \sim i.i.d.N(0, 1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	1.0752	0.9792	0.1	1	3	1.1996	1.0549	0.1	1	5	1.5413	1.2313
0.5	1	1	1.0496	0.9489	0.5	1	3	1.1941	1.0095	0.5	1	5	1.2848	1.0074
0.9	1	1	1.0884	0.9833	0.9	1	3	1.1595	0.9701	0.9	1	5	1.1993	0.9587
0.1	3	1	1.1341	0.7557	0.1	3	3	1.3001	0.7975	0.1	3	5	1.6770	0.9076
0.5	3	1	1.1205	0.7424	0.5	3	3	1.2475	0.7622	0.5	3	5	1.4143	0.7402
0.9	3	1	1.1138	0.7653	0.9	3	3	1.1794	0.7390	0.9	3	5	1.2438	0.7166
0.1	12	1	1.1493	0.6877	0.1	12	3	1.2829	0.5514	0.1	12	5	1.7512	0.5328
0.5	12	1	1.1687	0.7299	0.5	12	3	1.2071	0.5291	0.5	12	5	1.3540	0.4492
0.9	12	1	1.1505	0.7126	0.9	12	3	1.1386	0.6580	0.9	12	5	1.2135	0.5763

Part II.4: $(T, N) = (150, 150)$ and $\epsilon_t = 0.5\epsilon_{t-1} + \omega_t$ where $\omega_t \sim i.i.d.N(0, 1)$

N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D	N_1/N	h	r	$RMSFE^L$	L/D
0.1	1	1	1.0768	0.9576	0.1	1	3	1.2308	1.0608	0.1	1	5	1.7927	1.5644
0.5	1	1	1.0837	0.9712	0.5	1	3	1.1534	0.9837	0.5	1	5	1.2121	1.0207
0.9	1	1	1.0869	0.9709	0.9	1	3	1.1268	0.9629	0.9	1	5	1.1646	0.9903
0.1	3	1	1.1626	0.7721	0.1	3	3	1.3189	0.8404	0.1	3	5	2.0675	1.1331
0.5	3	1	1.1360	0.7663	0.5	3	3	1.1492	0.7358	0.5	3	5	1.2192	0.7106
0.9	3	1	1.1360	0.7663	0.9	3	3	1.1961	0.7589	0.9	3	5	1.2103	0.7322
0.1	12	1	1.1450	0.7022	0.1	12	3	1.3361	0.5976	0.1	12	5	2.1637	0.6903
0.5	12	1	1.1101	0.6740	0.5	12	3	1.1735	0.5779	0.5	12	5	1.2390	0.5081
0.9	12	1	1.1559	0.7017	0.9	12	3	1.1501	0.6654	0.9	12	5	1.2046	0.5901

Table 2: Performance of forecasts for 68 $I(1)$ U.S. monthly macroeconomic variables

	$h=1$	$h=3$	$h=6$	$h=12$	$h=24$
Period I: 1960 - 1983					
Min	0.8850	0.8294	0.7275	0.6349	0.5077
Max	1.1497	1.2243	1.2285	1.8387	1.5608
Mean	1.0022	0.9983	0.9850	1.0376	0.8026
10th percentile	0.9553	0.8927	0.8711	0.7434	0.5846
25th percentile	0.9734	0.9514	0.9224	0.8433	0.6452
50th percentile	0.9944	1.0129	0.9937	1.0043	0.7299
75th percentile	1.0264	1.0399	1.0454	1.2073	0.8803
90th percentile	1.0590	1.0879	1.1088	1.4195	1.1891
$\hat{Pr}(DM \geq 1.645)$	0.0882	0.0588	0.0735	0.1176	0.2353
$\hat{Pr}(DM \leq -1.645)$	0.0441	0.0441	0.0588	0.1471	0.0294
Period II: 1984 - June 2007					
Min	0.8626	0.7216	0.6100	0.4395	0.4196
Max	1.0810	1.1564	1.3892	1.6335	1.6966
Mean	0.9615	0.9204	0.8724	0.7602	0.7551
10th percentile	0.8986	0.7901	0.6784	0.5673	0.5497
25th percentile	0.9339	0.8418	0.7569	0.6072	0.6219
50th percentile	0.9634	0.9285	0.8464	0.7331	0.6994
75th percentile	0.9896	0.9870	0.9563	0.8536	0.8587
90th percentile	1.0141	1.0357	1.0991	0.9834	0.9921
$\hat{Pr}(DM \geq 1.645)$	0.2794	0.2647	0.2206	0.3382	0.3529
$\hat{Pr}(DM \leq -1.645)$	0.0000	0.0294	0.0294	0.0294	0.0000
Period III: July 2007 - October 2018					
Min	0.7511	0.5542	0.4640	0.3054	0.2094
Max	1.3211	1.3816	1.1155	1.2012	1.9306
Mean	0.9136	0.8067	0.7143	0.5916	0.6785
10th percentile	0.8058	0.6629	0.5189	0.4027	0.4236
25th percentile	0.8394	0.7180	0.5983	0.4449	0.4924
50th percentile	0.9039	0.7944	0.7004	0.5750	0.5578
75th percentile	0.9703	0.8772	0.8300	0.6979	0.7123
90th percentile	1.0349	0.9911	0.8928	0.7996	1.1123
$\hat{Pr}(DM \geq 1.645)$	0.5000	0.5294	0.5441	0.6324	0.4853
$\hat{Pr}(DM \leq -1.645)$	0.0147	0.0147	0.0000	0.0000	0.0441

Table 3: Performance of forecasts

1. Inflation rate

	$h = 1$			$h = 3$			$h = 6$		
	$RMSFE^L$	L/D	DM test	$RMSFE^L$	L/D	DM test	$RMSFE^L$	L/D	DM test
Period 1	2.4284	0.9485	1.6394 [0.1011]	2.5049	0.7348	2.9894 [0.0028]	2.5305	0.6519	6.5806 [0.0000]
Period 2	2.2572	1.0059	-0.3787 [0.7049]	2.4679	0.6696	2.8665 [0.0042]	2.4610	0.7039	2.5847 [0.0097]
Period 3	1.9366	0.9940	0.0606 [0.9517]	1.8352	0.6782	3.0700 [0.0021]	1.8531	0.6714	3.3499 [0.0008]
	$h = 12$			$h = 24$					
	$RMSFE^L$	L/D	DM test	$RMSFE^L$	L/D	DM test			
Period 1	2.5346	0.5624	5.0758 [0.0000]	2.4271	0.5440	3.1326 [0.0017]			
Period 2	2.3551	0.7698	2.0350 [0.0419]	2.6829	0.8532	1.2138 [0.2248]			
Period 3	1.7037	0.4250	4.1689 [0.0000]	1.7965	0.4540	2.0631 [0.0391]			

2. IP index

	$h = 1$			$h = 3$			$h = 6$		
	$RMSFE^L$	L/D	DM test	$RMSFE^L$	L/D	DM test	$RMSFE^L$	L/D	DM test
Period 1	0.0070	0.9888	0.1825 [0.8552]	0.0151	1.0302	-0.4973 [0.6190]	0.0236	1.1281	-10.6396 [0.0000]
Period 2	0.0043	0.9723	0.6965 [0.4861]	0.0080	0.9516	0.4907 [0.6237]	0.0129	0.8405	0.8327 [0.4050]
Period 3	0.0035	0.8402	2.4349 [0.0149]	0.0052	0.7390	3.9099 [0.0001]	0.0055	0.5420	2.8364 [0.0046]
	$h = 12$			$h = 24$					
	$RMSFE^L$	L/D	DM test	$RMSFE^L$	L/D	DM test			
Period 1	0.0336	1.2965	-1.3779 [0.1682]	0.0451	0.7259	1.1641 [0.2444]			
Period 2	0.0191	0.7645	1.1975 [0.2311]	0.0310	0.9299	0.3242 [0.7458]			
Period 3	0.0086	0.5769	2.9139 [0.0036]	0.0083	0.4572	1.7799 [0.0751]			

Note: (i) Periods I, II and III are, respectively, the pre-Great Moderation period (1960-1983), the Great Moderation period (1984-June, 2007), and the crisis and aftermath period (July, 2007-October, 2018).
(ii) $RMSFE^L$ and $RMSFE^D$ denote empirical root mean squared forecasting errors from $\hat{y}^L_{T+h}$ and $\hat{y}^D_{T+h}$, respectively, and $\frac{L}{D} = \frac{RMSFE^L}{RMSFE^D}$.
(iii) Values of the Diebold-Mariano test statistics and their corresponding p-values (in the parentheses) are reported under the column "DM test".

Table 4: Properties of forecasting residuals

1. Inflation rate

	$h = 1$			$h = 3$			$h = 6$		
	ADF test	LM test	ρ_ϵ	ADF test	LM test	ρ_ϵ	ADF test	LM test	ρ_ϵ
Period I	Reject	Reject	0.1916**	Reject	Do not reject	0.1811**	Reject	Reject	0.1892**
	-9.3307	0.4234	[0.0199]	-5.9950	0.2031	[0.0208]	-9.1932	0.2550	[0.0137]
Period II	Reject	Do not reject	0.0076	Reject	Reject	0.2521***	Reject	Do not reject	0.2662***
	-10.9859	0.2145	[0.9286]	-8.3356	0.4106	[0.0035]	-8.3136	0.2072	[0.0012]
Period III	Reject	Reject	0.2556**	Reject	Reject	0.2704**	Reject	Do not reject	0.3110***
	-5.9950	0.3101	[0.0404]	-5.7938	0.4533	[0.0192]	-5.3927	0.1772	[0.0048]
	$h = 12$			$h = 24$					
	ADF test	LM test	ρ_ϵ	ADF test	LM test	ρ_ϵ			
Period I	Reject	Reject	0.4162***	Reject	Do not reject	0.4440***			
	-6.9532	0.5017	[0.0000]	-6.3249	0.1118	[0.0000]			
Period II	Reject	Do not reject	0.1923**	Reject	Do not reject	0.1772**			
	-8.9769	0.1691	[0.0327]	-8.2008	0.1635	[0.0497]			
Period III	Reject	Do not reject	0.3507**	Reject	Do not reject	0.1017			
	-4.8383	0.1159	[0.0229]	-5.6359	0.1120	[0.5327]			

2. IP index

	$h = 1$			$h = 3$			$h = 6$		
	ADF test	LM test	ρ_ϵ	ADF test	LM test	ρ_ϵ	ADF test	LM test	ρ_ϵ
Period I	Reject	Reject	0.1840***	Reject	Reject	0.7664***	Do not reject	Reject	0.8480***
	-9.9283	0.7367	[0.0076]	-6.4097	0.4662	[0.0000]	-3.2011	0.3347	[0.0000]
Period II	Reject	Reject	0.1092	Do not reject	Reject	0.7065***	Do not reject	Reject	0.8674***
	-10.2613	0.4728	[0.1469]	-3.6431	0.6060	[0.0000]	-3.2988	0.5763	[0.0000]
Period III	Reject	Do not reject	0.1944	Do not reject	Do not reject	0.6375***	Do not reject	Do not reject	0.5702***
	-6.4097	0.1936	[0.1015]	-3.5573	0.1723	[0.0000]	-3.7703	0.1501	[0.0000]
	$h = 12$			$h = 24$					
	ADF test	LM test	ρ_ϵ	ADF test	LM test	ρ_ϵ			
Period I	Do not reject	Do not reject	0.8467***	Do not reject	Reject	0.8707***			
	-3.7802	0.1629	[0.0000]	-3.0207	0.3500	[0.0000]			
Period II	Do not reject	Reject	0.9194***	Do not reject	Do not reject	0.8974***			
	-4.5430	0.3858	[0.0000]	-2.2461	0.0901	[0.0000]			
Period III	Do not reject	Reject	0.5509***	Do not reject	Reject	0.8116***			
	-4.4332	0.4722	[0.0000]	-3.2391	0.6316	[0.0000]			

Note: Statistical decisions for the ADF and KPSS tests are made at the 5% significance level. Critical values for them were taken from Phillips and Ouliaris (1990) and Choi and Ahn (1995). Estimates of the AR(1) coefficient which are significant at the 10%, 5%. and 1% levels are marked by "*", "**", and "***", respectively.

Figure 1: Distributions of the relative efficiency measures $\frac{L}{D}$