Differencing versus Non-Differencing in Factor-Based Forecasting

In Choi* and Hanbat Jeong†

First Draft: July, 2018

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 of the two competing forecasts. Simulation results indicate that forecasting using non-differenced data performs better in terms of mean squared forecasting errors. The advantage of using non-differenced data is more pronounced when the forecasting horizon is long and the number of factors is large. This paper applies the two competing forecasts to U.S. inflation and finds that forecasts using non-differenced data outperform those using differenced data except for one-month forecasting horizon.

Keywords: nonstationary factors, factor-based forecasting, U.S. inflation rates

*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@osu.edu

1 Introduction

Factor models have been utilized for various purposes in economics and finance: (i) formulation of economic indicators, (ii) forecasting, (iii) policy analysis, (iv) instrumental variables estimation, and (v) modeling cross-sectional correlation. Breitung and Eickmeier (2006), Bai and Ng (2008), Stock and Watson (2011) and Breitung and Choi (2013) survey the literature on factor analysis. 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. It extracts estimated factors from a large dataset and employ 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 the differenced data are deemed to be stationary and the standard theory of inference can be applied for the differenced data. However, it is uncertain whether the common practice provides best forecasting performance. 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. Properties of the estimators of nonstationary factors and related model selection criteria are studied in Bai (2004). In addition, Choi (2017) suggests the generalized principal component estimator for nonstationary factors.

We expect three contributions through this paper. 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 employ $I(0)$ 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 from the approximate variances that forecasts based on differenced data can be advantageous only if we have many additional stationary variables for estimating factor spaces, and that their approximate variance increases as does the forecasting horizon. By contrast, the approximate variance of forecasts based on level data does not vary much with the forecasting horizon.

Second, we compare root mean squared forecasting errors of the two forecasts via Monte Carlo simulations. Simulation results indicate that the forecasting method using non-differenced data usually performs better in terms of root mean squared forecasting errors. The advantage of using non-differenced data is more pronounced when the forecasting horizon is long and the number of factors is large. A practical implication of these results is that non-differenced data should be preferred for long-run prediction. Using differenced data can be favorable only if the following conditions hold simultaneously: (i) a researcher can employ many $I(0)$ variables for estimating factor spaces, (ii) the number of factors is large, and (iii) the forecasting horizon is short.

Third, we apply the two competing forecasts, one using differenced data and the other using non-differenced ones, to CPI-based U.S. inflation rate and compare their performance based on ex-post forecasting errors. The data are borrowed from Stock and Watson (2005). Stock and Watson (1999) and Canova (2007) report that U.S. and U.K. inflation rates are difference-stationary processes. Since employing stationary variables has been conventional in the literature, differences of U.S. and U.K. inflation rates have often been forecasted. Except for one-month forecasting horizon, we find that forecasts using non-differenced data outperform those using differenced data. The Diebold-Mariano test results confirm that the better performance is statistically significant in most cases. Since only 6 or 7 common factors appear to be present at the dataset and 31 variables out of 132 are stationary, these results appear to support our theoretical and simulation results.

This paper proceeds as follows. Section 2 discusses motivation of this paper. Section 3 states the model, assumptions and preliminary results. Section 4 introduces main theoretical results of this paper. Section 5 reports simulation results for the two forecasts using non-differenced and differenced data. Section 6 compares the performance of the two forecasts for CPI-based U.S. inflation rate. Section 7 summarizes and provides further remarks. All proofs are in appendices.

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 $a = (a_1, \cdots, a_k)'$, $\|a\| = \left(a_1^2 + \cdots, a_k^2\right)^{\frac{1}{2}}$ 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 $\xrightarrow{p}$ and $\xrightarrow{d}$, respectively. The usual difference operator is denoted as Δ (i.e., $\Delta x_t = x_t - x_{t-1}$).

2 Motivation

Economists and policymakers have always regarded the inflation rate as one of the most important macroeconomic variables. The inflation rate is a measure of price inflation/deflation, which is usually defined by the annualized percentage change of the consumer price index (hereafter, CPI). Economists have studied on causes of inflation as well as effects of inflation on other economic variables. For a cause of inflation, there is a consensus that the inflation rate is crucially affected by the growth rate of money supply and the growth of the economy in the long run. It is called the quantity theory of inflation, which is based on Fisher's equation of exchange. In the short run, however, Keynesian economists argue that inflation can be affected by both supply and demand sides of an economy as well as various economic indicators such as interest rates, unemployment rate[1], wage elasticities and so on. According to this point of view, the inflation rate can be explained by various economic variables and might have the following statistical relationship: for each t,

$$y_t = f\left(X_{1t}, \cdots, X_{Nt}\right) + \epsilon_t \tag{1}$$

where y_t denotes the t^{th}-period inflation rate, X_{1t}, $\cdots$, X_{Nt} are N economic variables potentially affecting the inflation rate, $f(\cdot)$ is a function characterizing a relationship between X_{1t}, $\cdots$, X_{Nt} and y_t, and ϵ_t is a statistical error. Suppose $f(\cdot)$ takes a linear specification. When N is small, conventional multivariate time-series models can be used to explain the evolvement of y_t. However, if N becomes large, performance of the econometric model (1) may deteriorate due to the curse of dimensionality.

To resolve this issue, a researcher may employ one of the following two ways. First, one can take a penalizing procedure, which yields a manageable number of predictors. In this approach, shrinkage techniques are used to penalize the growing number of parameters. Examples are the LASSO (least absolute shrinkage and selection operator) estimator and its variants like the Ridge and elastic net estimators (cf. Bühlmann and van de Geer, 2011). Second, we may use factor models for X_{1t}, $\cdots$, X_{Nt} and use the estimated factors to explain y_t. Stock and Watson (1999, 2002) and Bai and Ng (2006) take this approach, which we also adopt in this paper. More specifically, $f\left(X_{1t}, \cdots, X_{Nt}\right)$ will be replaced by a function of unobserved, common factors.

Conventionally, all economic variables X_{1t}, $\cdots$, X_{Nt} and y_t are assumed to be stationary and the standard econometric theory can be applied under the factor-based approach. In the case of forecasting U.S. or U.K. inflation rates, however, the targeted variable y_t turns out to be $I(1)$ as shown in Stock and Watson (1999) and Canova (2007). We plotted in Figure 1 the monthly CPI-based U.S. inflation rate and its autocorrelation (AC) and partial autocorrelation (PAC) functions during the period 1960–2003 taken

[1]The Phillips curve describes the trade-off relationship between inflation and unemployment. Recently, this relationship turns out to be complicated, so a linear relationship might not capture the relationship between two variables (Hossfeld (2010)). However, various modifications of the Phillips curve give us evidence that the unemployment rate is still important to expect the inflation rate.

from Stock and Watson's (2005) dataset.

[Figure 1 here]

This figure shows that there is no noticeable trend in the data.[2] The AC function shows a slow decay, suggesting nonstationarity. For PAC, the first two and some lags (ninth, eleventh, twelfth, and thirteenth) are significant. It provides evidence for a long memory of the U.S. inflation rate. Moreover, the null hypothesis of a unit root for the U.S. inflation rate is not rejected at conventional significance levels (the value of the augmented Dicky-Fuller (ADF) test statistic is -2.3969 and its corresponding p-value is 0.1431).[3] After differencing the U.S. inflation rate, the value of the ADF test statistic and its corresponding p-value are respectively -14.1428 and 0.0000. This confirms that the U.S. inflation rate is a difference stationary process.

The dataset $(y_t, X_{1t}, \cdots, X_{Nt})$ usually consists of $I(1)$ and $I(0)$ variables.[4] In this paper, we suggest a forecasting method using only $I(1)$ variables if y_t is an $I(1)$ variable. We may have a relatively small number of variables in use if only $I(1)$ variables are used. However, there can be efficiency gains relative to the conventional forecasting method based on $I(0)$ variables since differencing the data may lead to the loss of information.

3 The model, assumptions and preliminary results

Suppose that the $I(1)$ variables $\{X_t\}$ is modeled as

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

where X_t is an $N_1 \times 1$ vector of observations, Λ is an $N_1 \times r$ matrix of factor-loadings, $F_t = [f_{t,1}, \cdots, f_{t,r}]'$ is an $r \times 1$ vector of latent factors and e_t is an idiosyncratic error 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 processes and F_0 is a random vector. We will introduce more specific assumptions on $\{F_t\}$ later. The number of factors r is assumed to be known throughout the

[2] Indeed, the coefficient of the linear time trend is insignificant when we estimate the $AR(1)$ model with constant and the time trend.

[3] The value of the KPSS test statistic is 0.5280, implying rejection of the null hypothesis of stationarity at the 5% significance level.

[4] In Stock and Watson's (2005) dataset, there are $I(0)$, $I(1)$, and $I(2)$ economic variables.

paper.[5] Methods to estimate r are introduced in Bai (2004). Note that N_1 is the number of $I(1)$ variables. Let N_0 denote the number of $I(0)$ variables so that the total number of variables becomes $N = N_0 + N_1$. We assume $\frac{N_1}{N} \to c$ as $N \to \infty$, where $c \in (0,1)$.

Model (2) can be represented in scalar notation as

$$X_{it} = \lambda_i' F_t + e_{it}, \ (i = 1, \cdots, N_1; \ t = 1, \cdots, T) \tag{3}$$

where λ_i' is the i-th row of matrix Λ. In matrix notation, model (2) is written as

$$X = F\Lambda' + e, \tag{4}$$

where $X = [X_1, \cdots, X_T]'$, $F = [F_1, \cdots, F_T]'$ and $e = [e_1, \cdots, e_T]'$.

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

$$\min_{\Lambda, F} \frac{1}{N_1 T} \sum_{i=1}^{N_1} \sum_{t=1}^{T} (X_{it} - \lambda_i' F_t)^2.$$

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 times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix XX'. The PCE of Λ is given by $\tilde{\Lambda}^L = \frac{1}{T^2} X' \tilde{F}$.

Alternatively, we may difference model (2) and combine the resulting model with stationary variables $\{x_t^s\}$ such that

$$\begin{bmatrix} \Delta X_t \\ x_t^s \end{bmatrix} = \begin{bmatrix} \Lambda \\ \Lambda^s \end{bmatrix} \Delta F_t + \begin{bmatrix} \Delta e_t \\ e_t^s \end{bmatrix} = \begin{bmatrix} \Lambda \\ \Lambda^s \end{bmatrix} u_t + \begin{bmatrix} \Delta e_t \\ e_t^s \end{bmatrix}, \ (t = 2, \cdots T) \tag{5}$$

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 disturbances for x_t^s.[6] One can estimate the space spanned by u_t using the principal component estimation method as in Bai (2003). This estimator is denoted as $\tilde{u}_t$. Model (5) is written in matrix notation as

$$\Delta X = u \begin{bmatrix} \Lambda \\ \Lambda^s \end{bmatrix}' + \Delta e,$$

[5] In our setting, r does not depend on N_0, N_1 and N.

[6] 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 do that since such practice brings $I(1)$ disturbances.

where $u = [u_2, \cdots, u_T]'$, $\Delta X = \left[(\Delta X_2', x_2^{s\prime})', \cdots, (\Delta X_T', x_T^{s\prime})'\right]'$ and Δe is defined similarly. An advantage of employing model (5) is that additional stationary variables can be used. However, we lose one period observations X_1 by differencing.

Now, we introduce some assumptions for the asymptotic distributions of the estimated factor spaces. These are taken from of Bai (2003, 2004) and reported here for completeness. In the below, the notation M stands for a finite positive constant, not depending on N and T.

Assumption 1 *(i) For some $\delta > 0$ and for all $t \le T$, $E\|u_t\|^{4+\delta} \le M$.*
(ii) As $T \to \infty$, $\frac{1}{T^2}\sum_{t=1}^{T} F_t F_t' \xrightarrow{d} \int_0^1 B_F(s)B_F'(s)ds$, where $B_F(s)$ is an r-dimensional Brownian motion with a positive definite covariance matrix $\Phi_F = \sum_{k=-\infty}^{\infty} E(u_1 u_k')$.
(iii) As $T \to \infty$, $\frac{\sum_{t=1}^{T} u_t u_t'}{T} \xrightarrow{p} \Sigma_u \ (> 0)$.
(iv) $E\|F_0\|^4 \le M$.

Assumption 2 *(i) For all i, $E\|\lambda_i\|^4 \le M$ and $E\|\lambda_i^s\|^4 \le M$.*
(ii) As N_0, $N_1 \to \infty$, $\frac{\Lambda'\Lambda}{N_1} \xrightarrow{p} \Sigma_\Lambda > 0$ and $\frac{\Lambda^{s\prime}\Lambda^s}{N_0} \xrightarrow{p} \Sigma_{\Lambda^s} > 0$ where Σ_Λ and Σ_{Λ^s} are $r \times r$ positive definite nonrandom matrices.

Assumption 3 *(i) $E(e_{it}) = 0$ and $E|e_{it}|^8 \le M$.*

(ii) Let $E\left(\frac{e_s' e_t}{N}\right) = \gamma_N(s,t)$. Then, $|\gamma_N(s,t)| \le M$ for all s and $\sum_{s=1}^{T} |\gamma_N(s,t)| \le M$ for each t.

(iii) Let $E(e_{it}e_{jt}) = \omega_{ij,t}$. Then, $|\omega_{ij,t}| \le |\omega_{ij}|$ for some ω_{ij} that satisfies the relation $\sum_{j=1}^{N} |\omega_{ij}| \le M$ for each i.

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

(v) For every (t,s), $E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[e_{is}e_{it} - E(e_{is}e_{it})\right]\right|^4 < \infty$.

Assumption 4 *Assumption 3 holds when e_{it} is replaced with Δe_{it} or e_{it}^s.*

Assumption 5 *(i) $\{\lambda_i, \lambda_i^s\}$, $\{u_t\}$ and $\{e_{it}, e_{it}^s\}$ are mutually independent.*
(ii) The eigenvalues of the matrix $(c\Sigma_\Lambda + (1-c)\Sigma_{\Lambda^s})^{\frac{1}{2}} \Sigma_u (c\Sigma_\Lambda + (1-c)\Sigma_{\Lambda^s})^{\frac{1}{2}}$ are distinct.
(iii) The eigenvalues of the matrix $\Sigma_\Lambda^{\frac{1}{2}} \int_0^1 B_F(r)B_F'(r)dr \Sigma_\Lambda^{\frac{1}{2}}$ are distinct almost surely.
(iv) For each t, as $N \to \infty$,

$$(a)\ \frac{1}{\sqrt{N_1}}\sum_{i=1}^{N_1} \lambda_i e_{it} \xrightarrow{d} N(0, \Sigma_t);$$

$$(b) \ \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{N_1} \lambda_i \Delta e_{it} + \sum_{i=N_1+1}^{N} \lambda_i^s e_{it}^s \right) \xrightarrow{d} N(0, \Xi_t),$$

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

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

The Assumption 1, 2, 3, 4 and 5 are now standard for the approximate factor model. Assumption 1 (ii) follows from the functional central limit theorem (cf. Phillips and Durlauf, 1986). The positive-definiteness of Φ_F rules out cointegration among the elements of F_t. Assumption 2 is about the factor loading matrix. It is essentially the law of large numbers for $\{\lambda_i\}$. Assumption 3 controls the degree of serial and cross-sectional dependency in the the idiosyncratic errors. Assumption 4 is the same as Assumption 3 with e_{it} replaced by Δe_{it} or e_{it}^s. Assumption 5 is a technical assumption required for the asymptotic distributions of the principal component estimators. The reader is referred to Bai (2003, 2004) for further explanation on these assumptions.

The following lemma reports the asymptotic distributions of estimated factor spaces $\tilde{F}_t$ and $\tilde{u}_t$. Part (i) is from Lemma 2 and Corollary 1 of Bai (2004). Part (ii) is a slight modification of Theorem 1 of Bai (2003). Its proof is not worth reporting here. The results and notation of the following lemma will be used in the next section.

Lemma 1 *Suppose that Assumptions 1, 2, 3, 4 and 5 hold.*
(i) (a) If $\frac{N_1}{T^3} \to 0$, *for each* t,

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

where $H_L = \left(\frac{\Lambda' \Lambda}{N_1} \right) \left(\frac{F' \tilde{F}}{T^2} \right) \left(V_{N_1 T}^L \right)^{-1}$, $V_{N_1 T}^L$ *is an* $r \times r$ *diagonal matrix consisting of the* r *largest eigenvalues of* $\frac{XX'}{N_1 T^2}$ *in descending order,* $V^L = diag \left[v_1^L, \cdots, v_r^L \right]$, $v_1^L > \cdots > v_r^L > 0$ *are the eigenvalues of* $\Sigma_\Lambda^{1/2} \int_0^1 B_F(r) B_F'(r) dr \Sigma_\Lambda^{1/2}$, Υ^L *is the corresponding eigenvalue matrix such that* $\Upsilon^{L\prime} \Upsilon^L = I_r$ *and* $Q^L = \left(V^L \right)^{1/2} \Upsilon^{L\prime} \Sigma_\Lambda^{-1/2}$.
(b) If $\liminf \frac{N_1}{T^3} \geq \delta > 0$, $\tilde{F}_t - H_L' F_t = O_p(T^{-3/2})$.
(ii) (a) If $\frac{\sqrt{N}}{T} \to 0$, *for each* t,

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

where $H_D = \left(\frac{\Lambda'\Lambda + \Lambda^{s\prime}\Lambda^s}{N}\right)\left(\frac{u'\tilde{u}}{T}\right)\left(V_{NT}^D\right)^{-1}$, V_{NT}^D *is an* $r \times r$ *diagonal matrix consisting of the* r *largest eigenvalues of* $\frac{\Delta X \Delta X'}{NT}$ *in descending order,* $V^D = diag\left[v_1^D, \cdots, v_r^D\right]$, $v_1^D > \cdots > v_r^D > 0$ *are the eigenvalues of* $\left(c\Sigma_\Lambda + (1-c)\,\Sigma_{\Lambda^s}\right)^{\frac{1}{2}} \Sigma_u \left(c\Sigma_\Lambda + (1-c)\,\Sigma_{\Lambda^s}\right)^{\frac{1}{2}}$, Υ^D *is the corresponding eigenvalue matrix such that* $\Upsilon^{D\prime}\Upsilon^D = I_r$ *and* $Q^D = \left(V^D\right)^{\frac{1}{2}} \Upsilon^{D\prime}\Sigma_\Lambda^{-1/2}$.

(b) If $\liminf \frac{\sqrt{N}}{T} \geq \tau > 0$, $\tilde{u}_t - H_D' u_t = O_p(\frac{1}{T})$ *for each* t.

This lemma shows that the factor spaces can be estimated consistently whether the data are differenced or not. The convergence rates for both the cases are essentially the same ($\sqrt{N}$) according to parts (i) (a) and (ii) (a). However, parts (i) (b) and (ii) (b) show that the estimator of the factor spaces using non-differenced data can converge faster than that using differenced data. But it is not straightforward to compare the asymptotic variances of $\tilde{F}_t$ and $\tilde{u}_t$ when they have the same rate of convergence.

4 Factor based forecasting models

This section derives approximate variances of the forecasting errors from the forecasting regressions using $\tilde{F}_t$ and $\tilde{u}_t$. First, consider the regression equation using the non-differenced data

$$y_{t+h} = \mu + \alpha' F_t + \beta' Z_t + \epsilon_{t+h}, \; (t = 1, \cdots, T-h), \tag{6}$$

where h is the forecasting horizon, F_t is a vector of nonstationary factors from model (2), Z_t is a $K \times 1$ vector of observable variables that may include a linear time trend and $I(1)$ variables and ϵ_{t+h} is an $I(0)$ error term.

Estimating the parameters μ, α and β from model (6), we can forecast y_{T+h}. Since F_t is not observed, we use $\tilde{F}_t$ to estimate model (6). Then, using the notation of Lemma 1, the forecasting equation with $\tilde{F}_t$ can be written as

$$\begin{aligned} y_{t+h} &= \mu + \alpha' H_L'^{-1}\tilde{F}_t + \beta' Z_t + \epsilon_{t+h} + \alpha' H_L'^{-1}\left(H_L' F_t - \tilde{F}_t\right) \\ &= \delta^{L\prime}\tilde{L}_t + \epsilon_{t+h} + \alpha' H_L'^{-1}\left(H_L' F_t - \tilde{F}_t\right) \end{aligned} \tag{7}$$

where $\delta^L = \left(\mu, \alpha' H_L'^{-1}, \beta'\right)'$ and $\tilde{L}_t = \left(1, \tilde{F}_t', Z_t'\right)'$. Model (7) provides the OLS estimator $\hat{\delta}^L$ and the forecast of y_{T+h} is $\hat{y}_{T+h|T}^L = \hat{\delta}^{L\prime}\tilde{L}_T$. Hereafter, we denote this forecast as LF.

We can also forecast the target variable y_{T+h} using differenced data. In this case, letting $w_{t+h} = \Delta y_{t+h}$, the forecasting regression equation based on differenced data is

$$\begin{aligned} w_{t+h} &= \mu + \alpha' \Delta F_t + \beta' \Delta Z_t + \Delta \epsilon_{t+h}, \; (t = 2, \cdots, T-h) \\ &= \mu + \alpha' u_t + \beta' \Delta Z_t + \Delta \epsilon_{t+h}. \end{aligned} \tag{8}$$

If (6) is the true data generating process, the constant term in equation (8) should be zero. But the constant term will be retained here following the convention of a linear regression. Because u_t is not observable, we need $\tilde{u}_t$ to estimate model (8) instead of u_t. The forecasting equation using $\tilde{u}_t$ is written as

$$\begin{aligned} w_{t+h} &= \mu + \alpha' H_D'^{-1} \tilde{u}_t + \beta' \Delta Z_t + \Delta \epsilon_{t+h} + \alpha' H_D'^{-1} \left(H_D' u_t - \tilde{u}_t \right) \\ &= \delta^{D\prime} \tilde{P}_t + \Delta \epsilon_{t+h} + \alpha' H_D'^{-1} \left(H_D' u_t - \tilde{u}_t \right), \end{aligned} \tag{9}$$

where $\tilde{P}_t = (1, \tilde{u}_t', \Delta Z_t')'$. Using model (9), 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 $\hat{y}^D_{T+h|T} = \sum_{m=1}^{h} \hat{w}_{T+m|T} + y_T$, which is denoted as DF hereafter.

We need to make the following assumption to derive the approximate forecasting error variances.

Assumption 6 *Let $L_t = (1, F_t', Z_t')'$ and $\Delta L_t = (1, \Delta F_t', \Delta Z_t')'$. For any $h = 1, 2, \cdots$, the following holds jointly.*

(i) $\{F_t, Z_t, \epsilon_t\}$ and $\{e_{it}, e_{it}^s\}$ are independent for all i.

(ii) $\begin{pmatrix} D_T^{-1} \sum_{t=1}^{T-h} L_t L_t' D_T^{-1} \\ D_T^{-1} \sum_{t=1}^{T-h} L_t \varepsilon_{t+h} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \Sigma_L \\ \Sigma_{\varepsilon L}^{1/2} \times N(0, I) \end{pmatrix}$ where D_T is a diagonal matrix whose elements are functions of T, $\Sigma_L > 0$ and $\Sigma_{\varepsilon L} > 0$ with probability one, and $\Sigma_{\varepsilon L}^{1/2}$ is independent of $N(0, I)$.

(iii) $\frac{1}{T} \sum_{t=2}^{T-h} \Delta L_t \Delta L_t' \xrightarrow{p} \Sigma_{\Delta L} > 0$.

(iv) $\frac{1}{\sqrt{T}} \sum_{t=2}^{T-h} \Delta L_t \Delta \epsilon_{t+h} \xrightarrow{d} N(0, \Sigma_{\Delta \epsilon L})$,
where $\Sigma_{\Delta \epsilon L} = \lim_{T \to \infty} \frac{1}{T} E \left(\sum_{t=2}^{T} \Delta L_t \Delta \epsilon_{t+h} \right) \left(\sum_{t=2}^{T} \Delta L_t \Delta \epsilon_{t+h} \right)'$.

Part (ii) 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,

$$\Sigma_L = \begin{bmatrix} 1 & \int_0^1 B_F'(r) dr \\ \int_0^1 B_F(r) dr & \left(\int_0^1 B_F(r) B_F'(r) dr \right) \end{bmatrix} \oplus \Sigma_z,$$

where $\Sigma_z = \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'$ 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

$$\Sigma_L = \begin{bmatrix} 1 & \int_0^1 B_{FZ}'(r)dr \\ \int_0^1 B_{FZ}(r)dr & \left(\int_0^1 B_{FZ}(r)B_{FZ}'(r)dr\right) \end{bmatrix},$$

where $B_{FZ}(r) = (B_F'(r), B_Z'(r))'$ and $B_Z(r)$ is the weak limit of $\frac{1}{\sqrt{T}}Z_{[Tr]}$. Part (ii) also assumes the limiting distribution of $D_T^{-1}\sum_{t=1}^{T} L_t\varepsilon_{t+h}$. If $\{F_t\}$ are independent of $\{\varepsilon_t\}$ and if $\{Z_t,\ \varepsilon_{t+h}\}$ is a stationary and ergodic process, we have with $D_T = diag[\sqrt{T}, T, ..., T, \sqrt{T}, ..., \sqrt{T}]$

$$\Sigma_{\varepsilon L} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \int_0^1 B_F'(r)dr \\ \int_0^1 B_F(r)dr & \left(\int_0^1 B_F(r)B_F'(r)dr\right) \end{bmatrix} \oplus \lim_{T\to\infty}\frac{1}{T}E(\sum_{t=1}^{T-h} Z_t\varepsilon_{t+h})(\sum_{t=1}^{T-h} Z_t\varepsilon_{t+h})',$$

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

$$\Sigma_{\varepsilon L} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \int_0^1 B_{FZ}'(r)dr \\ \int_0^1 B_{FZ}(r)dr & \left(\int_0^1 B_{FZ}(r)B_{FZ}'(r)dr\right) \end{bmatrix},$$

with $D_T = diag[\sqrt{T}, T, ..., T]$. Other cases of $\{Z_t\}$ containing a linear time trend can also be considered in a similar manner. Part (iii) is the law of large numbers for $\frac{1}{T}\sum_{t=2}^{T-h}\Delta L_t \Delta L_t'$. Part (iv) is a central limit theorem for $\frac{1}{\sqrt{T}}\sum_{t=2}^{T-h}\Delta L_t \Delta\epsilon_{t+h}$.

The asymptotic distributions of OLS estimators $\hat{\delta}^L$ and $\hat{\delta}^D$ are reported in the following lemma. This lemma will be used to derive the approximate forecasting error variances.

Lemma 2 *Suppose that Assumptions 1, 2, 3, 4, 5 and 6 hold. Let* $\Psi_L = 1 \oplus \left(V^L\right)^{-1/2}\Upsilon^{L\prime}\Sigma_\Lambda^{1/2} \oplus I_K$ *and* $\Psi_D = 1 \oplus \left(V^D\right)^{-1/2}\Upsilon^{D\prime}\Sigma_\Lambda^{1/2} \oplus I_K$.
(i) If $\frac{T}{N_1} \to 0$ *and* $\frac{N_1}{T^3} \to 0$ *or if* $\liminf \frac{N_1}{T^3} \geq \delta > 0$,

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

If $\{F_t, Z_t\}$ *and* $\{\epsilon_t\}$ *are independent,* $\Psi_L'^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}^{1/2}$ *is independent of* $N(0, I)$.
(ii) If $\frac{T}{N} \to 0$ *and* $\frac{\sqrt{N}}{T} \to 0$ *or if* $\liminf \frac{\sqrt{N}}{T} \geq \tau > 0$,

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

Now, we derive approximate variances of forecasting errors. The forecasting error from LF is written as

$$\hat{y}_{T+h|T}^L - y_{T+h} = \left(\hat{\delta}^L - \delta^L\right)'\tilde{L}_T + \alpha' H_L'^{-1}\left(\tilde{F}_T - H_L' F_T\right) - \epsilon_{T+h}, \qquad (10)$$

and that from DF is

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

The approximate variances of the forecasting errors are reported in the following theorem.

Theorem 3 *Suppose that Assumptions 1, 2, 3, 5 and 6 hold. In addition, we assume* $Var(\epsilon_t) = \sigma^2_\epsilon$ *for all t;* $E(\epsilon_t \epsilon_s) = 0$ *for* $t \neq s$*; and that* $E\left(\epsilon_{t+h}\epsilon_{T+h} | L_1, \cdots, L_{T-h}\right) = 0$ *for* $t = 1, \cdots, T-h$*. Let* $\widehat{Var}\left(\hat{y}^L_{T+h|T} - y_{T+h}\right)$ *and* $\widehat{Var}\left(\hat{y}^D_{T+h|T} - y_{T+h}\right)$ *be the approximate variances of the forecasting errors given* $\{L_t\}_{t=1}^T$.

(i) If $\frac{T}{N_1} \to 0$ *and* $\frac{N_1}{T^3} \to 0$,

$$\widehat{Var}\left(\hat{y}^L_{T+h|T} - y_{T+h}\right) = \underbrace{L_T' D_T^{-1} \Sigma_L^{-1} \Sigma_{\epsilon L} \Sigma_L^{-1} D_T^{-1} L_T}_{\textit{forecasting regression part}} + \underbrace{\frac{1}{N_1} \alpha' \left(\Sigma_\Lambda^{-1} \Sigma_T \Sigma_\Lambda^{-1} \right) \alpha}_{\textit{factor estimation part}} + \sigma^2_\epsilon \tag{11}$$

(ii) If $\frac{T}{N} \to 0$ *and* $\frac{\sqrt{N}}{T} \to 0$,

$$\begin{aligned}
\widehat{Var}\left(\hat{y}^D_{T+h|T} - y_{T+h}\right) &= \underbrace{\frac{1}{T} \left(\sum_{m=1}^{h} \Delta L_{T+m-h} \right)' \Sigma_{\Delta L}^{-1} \Sigma_{\Delta\epsilon L} \Sigma_{\Delta L}^{-1} \left(\sum_{m=1}^{h} \Delta L_{T+m-h} \right)}_{\textit{forecasting regression part}} \qquad (12) \\
&\quad \underbrace{+ \frac{1}{N} \alpha' \left(\Sigma_\Lambda^{-1} \sum_{m=1}^{h} \Xi_{T+m-h} \Sigma_\Lambda^{-1} \right) \alpha}_{\textit{factor estimation part}} + 2\sigma^2_\epsilon.
\end{aligned}$$

Some interpretations of this theorem are in order. First, we have $\frac{\widehat{Var}\left(\hat{y}^L_{T+h|T} - y_{T+h}\right)}{\widehat{Var}\left(\hat{y}^D_{T+h|T} - y_{T+h}\right)} \xrightarrow{p} \frac{1}{2}$ as $N, T \to \infty$, which implies that LF is more accurate in the limit than DF, although admittedly this might be an oversimplification.

Second, 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 (12) is larger than that of (11). 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 (11) becomes. In other words, LF tends to be accurate when the majority of the variables is $I(1)$.

Third, it is not straightforward to compare the first terms on the right-hand-sides of equations (11) and (12). But some terms of the matrix $D_T^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}\Sigma_L^{-1}D_T^{-1}$ are $O_p(\frac{1}{T^2})$, while every term of the matrix $\frac{1}{T}\left(\sum_{m=1}^{h}\Delta L_{T+m-h}\right)'\Sigma_{\Delta L}^{-1}\Sigma_{\Delta\epsilon L}\Sigma_{\Delta L}^{-1}\left(\sum_{m=1}^{h}\Delta L_{T+m-h}\right)$ is $O_p(\frac{1}{T})$. Thus, in reality, the first term on the right-hand-side of equation (11) may tend to be smaller than that of (12).

Fourth, the forecasting horizon h affects the forecasting error variance of DF, but not that of LF. We may infer from this that the forecasting error variance of DF grows with h, while that of LF does not. This implies that the factor-based forecasts using non-differenced data may have an advantage in long-run forecasting. In fact, this is what we would observe in the simulation results of the next section.

5 Simulation

This section compares the forecasting performance of LF and DF by simulation. The data for $\{X_{it}\}$ were generated by the following data generating process (DGP):

$$\begin{aligned}
X_{it} &= \lambda_i' F_t + \sqrt{\frac{r(1-\rho_i^2)}{\sigma_{\xi i}^2}}e_{it} \text{ (for } i = 1,\cdots,N_1); \qquad (13)\\
x_{it}^s &= \lambda_i^{s\prime} u_t + \sqrt{\frac{r(1-\rho_i^2)}{\sigma_{\xi i}^2}}e_{it} \text{ (for } i = N_1+1,\cdots,N);\\
\lambda_i &\backsim i.i.d.\ N(0,I_r) \text{ (for } i = 1,\cdots,N_1);\ \lambda_i^s \backsim i.i.d.\ N(0,I_r) \text{ (for } i = N_1+1,\cdots,N)\\
F_t &= F_{t-1} + u_t,\ u_t \backsim i.i.d.\ N(0,I_r) \text{ and } u_0 = 0;\\
e_{it} &= \rho_i e_{i,t-1} + \sigma_{\xi i}\xi_{it},\ \xi_{it} \backsim i.i.d.\ N(0,1) \text{ for } i = 1,\cdots,N,
\end{aligned}$$

where $\{\rho_i\}$ was taken from $U[0.3, 0.8]$ and $\{\sigma_{\xi i}^2\}$ from $U[1, 3]$. For $\{F_t\}$ and $\{e_{it}\}$, we generated $T + 60$ observations for each i and employed the last T series. Since

$$Var\left(\lambda_i' F_t\right) = tr$$

and

$$Var\left(\sqrt{\frac{r(1-\rho_i^2)}{\sigma_{\xi i}^2}}e_{it}\right) = \frac{r(1-\rho_i^2)}{\sigma_{\xi i}^2}\frac{\sigma_{\xi i}^2}{(1-\rho_i^2)} = r,$$

the signal-to-noise ratio is t for all the parameter values. For the sample size, we consider $(T, N) = (50, 80)$, $(50, 150)$, $(150, 80)$ and $(150, 150)$ with five cases of $\frac{N_1}{N}$ $(= \frac{1}{10}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}$ and $\frac{9}{10})$. For the number of factors, we set r to be 1, 3 and 5.

Next, the forecasting equation in levels is generated by using $\{F_t\}$ from DGP (13):

$$y_{t+h} = \mu + \alpha' F_t + \beta y_t + \epsilon_{t+h}, \ (t = 1, \cdots, T-h). \tag{14}$$

Throughout the simulation, we set $(\mu, \alpha', \beta) = \left(0.5, \underbrace{1, \cdots, 1}_{r \text{ times}}, 0.2\right)$ and $\epsilon_t \backsim i.i.d.\ N(0,1)$. For the forecasting horizon h, we consider $h = 1$, 3, 6, and 12. The target variable y_{t+h} can also be forecasted by using the model

$$\Delta y_{t+m} = \alpha' u_t + \beta \Delta y_t + \Delta \epsilon_{t+h}, \ (t = 1, \cdots, T-h; m = 1, \cdots, h). \tag{15}$$

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 mean squared forecasting errors of the two forecasts. The number of iterations for the calculation of the empirical mean squared forecasting errors is set to be 2,000. Let $RMSE^L_h$ and $RMSE^D_h$ denote respectively empirical root mean squared forecasting errors from models (6) and (8).[7] To illustrate relative efficiency of the two forecasts, we also report the ratio of $RMSE$s defined by $\frac{L}{D} = \frac{RMSE^L_h}{RMSE^D_h}$. Results in Table 1 are summarized as follows.

[Table 1 here]

(a) For almost all simulation environments, LF performs better than DF. Only when $r = 5$, $\frac{N_1}{N} = 0.1$ and $h = 1$, DF performs better than LF. These confirm the large-sample efficiency advantage of LF reported in Theorem 3. Moreover, the ratio of root mean squared forecasting errors $\frac{L}{D}$ is close to $\frac{1}{2}$ when T and N become large. This verifies the result $\frac{\widehat{Var}\left(\hat{y}^L_{T+h|T} - y_{T+h}\right)}{\widehat{Var}\left(\hat{y}^D_{T+h|T} - y_{T+h}\right)} \xrightarrow{p} \frac{1}{2}$ as $N, T \to \infty$ that is inferred from Theorem 3.

[7] $RMSE\left(\hat{y}^L_{T+h|T}\right)$ is defined as

$$RMSE\left(\hat{y}^L_{T+h|T}\right) = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(\hat{y}^L_{T+h|T,i} - y_{T+h,i}\right)^2},$$

where $\hat{y}^L_{T+h|T,i}$ is the forecast of $y_{T+h,i}$ in the i-th iteration and I is the number of iteration. $RMSE\left(\hat{y}^D_{T+h|T}\right)$ is defined similarly.

(b) Both $RMSE_h^L$ and $RMSE_h^D$ 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.

(c) As the forecasting horizon h increases, so does $RMSE_h^D$ while $RMSE_h^L$ behaves stably. We observe from $\frac{L}{D}$ that the relative efficiency gain of LF increases as does h. This indicates the advantage of LF in long-run forecasting.

(d) Even for very small $\frac{N_1}{N}$ (i.e., $\frac{N_1}{N} = 0.1$), LF usually performs better than DF. This result implies that using DF with many additional $I(0)$ variables is not recommended.

(e) $RMSE_h^L$ 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 $RMSE_h^L$ and $RMSE_h^D$ increase as does the true number of factors r. The relative efficiency gain of LF (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 LF corresponding to increasing $\frac{N_1}{N}$ becomes clearer under large r. This indicates that using LF has an advantage in case of large r.

In a nutshell, under DGPs (13), (14) and (15), LF tends to perform substantially better than DF. In the presence of a large numbers of $I(0)$ variables, DF has an advantage relative to LF at a short forecasting horizon

6 An empirical application

This section shows how to apply our prediction methods to a real dataset. We compare the performance of the two predictors $\hat{y}^L_{T+h|T}$ and $\hat{y}^D_{T+h|T}$ for the CPI-based U.S. inflation rate. For this, Stock and Watson's (2005) dataset containing 132 macroeconomic variables (i.e., $N = 132$) is used. The 132 monthly time series are available from January 1960 to December 2003. Based on the transformation rules given by Stock and Watson (2005), we categorize the 132 variables into three groups: 31 $I(2)$ variables, 73 $I(1)$ variables and 28 $I(0)$ variables. For the forecasting equation in differences, we use 132 transformed time series. The transformation rules introduced by Stock and Watson (2005) are followed. For the forecasting equation in levels, the 73 $I(1)$ variables and the 31 $I(2)$ variables are employed. The $I(2)$ variables are differenced once. That is, $N_1 = 101$ and $N_0 = 31$ which means that $\frac{N_1}{N}$ is 0.7652. We consider the initial sampling period from January 1960 to December 1977.[8] The reason why we take that initial sample is to compare forecasts from our model specifications with the predicted inflation rates by the University of Michigan Survey (which are available from January 1978).[9] Numbers of factors are estimated by well-known information criteria for factor models: (i) for the non-differenced data, Bai's (2004) IPC_1 and IPC_2 criteria suggest 6 nonstationary factors, and (ii) for the differenced data, Bai and Ng's (2002) IC_2 criterion selects 7 factors.

For the observable regressors $\{Z_t\}$, four kinds of variables are considered: (i) constant, (ii) past inflation rates[10], and (iii) the unemployment rate, and (iv) the term spread.[11] We do not include the linear time trend since we verify that there is no time trend in the U.S. inflation rate (see Section 2). The unemployment rate is often used to forecast inflation rates because of their stable relationship predicted by the Phillips curve. However, we observe a weak positive correlation between the CPI-based inflation rate and the unemployment rates (0.0872).[12] Although we do not find a proper negative statistical relationship and some articles reject the stable relationship empirically (cf. Atkeson and Ohanian, 2001; Fisher, Liu and Zhou, 2002; Ang, Bekaert and Wei, 2007; Stock and Watson, 2008a; 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 indicate 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 the 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. We observe a negative correlation

[8] When $h = 1$ is chosen, for example, $\hat{y}^L_{Jan1978|Dec1977}$ and $\hat{y}^D_{Jan1978|Dec1977}$ are firstly computed. After that, $\hat{y}^L_{Feb1978|Jan1978}$ and $\hat{y}^D_{Feb1978|Jan1978}$ are computed based on the data from January 1960 to January 1978.

[9] This monthly forecast is one of the most famous inflation predictions, and regularly reported in the Federal Reserve Bank's website (https://fred.stlouisfed.org/series/MICH). The data comes from University of Michigan's monthly survey of consumers' inflation expectations for the next 12 months. The reported monthly inflation forecast is the median of the collected individual forecasts. Strictly speaking, the forecast does not correspond to any particular forecasting horizon considered in this paper. For more information, see http://www.sca.isr.umich.edu/.

[10] For each forecasting exercise, the number of lags is selected via BIC. The maximum lags is set to be 20.

[11] For DF, Z_t are treated to be stationary.

[12] The correlation coefficient between $I(1)$ and $I(0)$ variables does not allow conventional interpretations. But we report it here as a basic statistic.

[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.

between the inflation rate and the term spread. For comparison, we report the prediction results from the four specifications for the explanatory variables: (i) only past inflation rates, (ii) past inflation rates and the unemployment rate, (iii) past inflation rates and the term spread, and (iv) past inflation rates with both the variables. For the forecasting horizon, we consider $h = 1, 3, 6, 12$ and 24.

Besides comparing forecasting performance, we test whether the two forecasts have statistically same prediction accuracy. For this, we utilize the Diebold-Mariano (hereafter, DM) test.[14] Consider two loss functions for the test: (i) $L_1(\xi_t) = \xi_t^2$ and (ii) $L_2(\xi_t) = |\xi_t|$. Then, the loss differential between the two forecasts is defined by

$$d_t = L(\xi_t^D) - L(\xi_t^L)$$

where $\xi_t^L = y_{t|t-h}^L - y_t$ and $\xi_t^D = y_{t|t-h}^D - y_t$.[15] The null hypothesis of the DM test is

$$H_0 : E(d_t) = 0 \text{ for all } t \tag{16}$$

and the alternative hypothesis is

$$H_1 : E(d_t) \neq 0 \text{ for some } t.$$

The two forecasts show statistically equal performance when the loss differential d_t has zero expectation for every t.

The DM statistic is defined by

$$DM = \frac{\bar{d}}{\sqrt{\frac{\hat{\gamma}_d(0)+2\sum_{k=1}^{h-1}\hat{\gamma}_d(k)}{J}}}$$

where $\bar{d} = \frac{1}{J}\sum_{t=1}^{J} d_t$ and $\hat{\gamma}_d(k) = \frac{1}{J}\sum_{t=|k|+1}^{T}\left(d_t - \bar{d}\right)\left(d_{t-|k|} - \bar{d}\right)$. To get improved results in small samples, we also employ the modified version of the DM test (hereafter, $HLN - DM$) suggested by Harvey, Leybourne, and Newbold (1997). The modified DM statistic is defined by

$$HLN - DM = \sqrt{\frac{J + 1 - 2h + h(h-1)/J}{J}} DM$$

where J denotes the number of forecasts.

To compare forecasting performance of LF and DF, we employ two criteria: the root mean squared

[14] The reader is referred to Diebold and Mariano (2002) and Diebold (2016) for more information on the DM test.

[15] We omit the subscript showing a loss function type since this definition holds for any loss function.

and mean absolute errors ($RMSE$ and MAE, respectively).[16] Table 2 reports results of the out-of-sample forecasting exercises using two equations (6) and (8). Values of the DM, $HLN-DM$ test statistics and corresponding p-values (in the parentheses), and the relative forecasting performance via the ratios of $RMSE$ and MAE (i.e., $\frac{RMSE_h^L}{RMSE_h^D}$ and $\frac{MAE_h^L}{MAE_h^D}$) are also presented there.

[Table 2 here]

The results in Table 2 are summarized below:

(a) The $RMSE$ and MAE of LF and DF increase when the forecasting horizon h increases. However, LF's performance does not vary much along with h compared to that of DF.

(b) At $h = 1$, DF with the specification (iv) performs better than the rest. Since both the corresponding DM and $HLN-DM$ test statistics are significant at the 1% level and take negative values, the better performance of DF is statistically confirmed. When $h = 3,$ 6 and 12, LF with the specification (iv) shows the best performance in terms of $RMSE$ and MAE, and both the corresponding DM and $HLN-DM$ test statistics are significant at the conventional significance levels. If $h = 24$, however, LF with specification (ii) shows the best performance and its better performance is statistically confirmed at the 5% significance level except when $L_2(\xi_t)$ is used.

(c) Except for the case of $h = 24$, both the unemployment rate and the term spread help to increase accuracy of LF and DF.

In summary, LF performs better than DF except at $h = 1$. This result is in accordance with the conclusion of our simulation study which states that LF tends to perform substantially better than DF and that DF has an advantage relative to LF at $h = 1$ in the presence of large numbers of factors and $I(0)$ variables.

[Figure 2 here]

Figure 2 shows the CPI-based inflation rate as well as the predicted inflation rate by the University of

[16] The RMSE and MAE are defined for $X = L, D$ as

$$\begin{aligned} RMSE_h^X &= \sqrt{\frac{1}{J}\sum_{j=0}^{J-1}\left(\hat{y}^X_{T+h+j|T} - y_{T+h+j}\right)^2} \\ MAE_h^X &= \frac{1}{J}\sum_{j=0}^{J-1}\left|\hat{y}^X_{T+h+j|T} - y_{T+h+j}\right|. \end{aligned}$$

Michigan Survey and our two forecasts at $h = 6$. Note that the survey-based forecasts show the central tendency of consumers' inflation expectations during the next 12 months (not after the 12 months). Since our forecasting models exactly specify the forecasting horizon h, we choose $h = 6$ as a middle value in order to compare our forecasts with the survey-based forecasts. On the y-axis, we also report histograms showing distributions of the U.S. inflation rate as well as its forecasts. Compared to the realized inflation rate, its forecasts tend to have small dispersions. Moreover, the survey-based forecasts show that consumers' expectations for price changes are very stable over time. Note that the $RMSE$ and MAE of the predicted inflation rates by the University of Michigan Survey are respectively 2.6546 and 1.9197 and those of our best predictor LF are respectively 2.5487 and 1.8461 at $h = 6$. This shows that our best predictor outperforms that of the University of Michigan Survey at $h = 6$.

7 Conclusion and further remarks

We have investigated the 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. We have also conducted simulations to compare the mean squared forecasting errors of the two competing forecasts. Simulation results indicate that the forecasting using non-differenced data usually performs better in terms of mean squared forecasting errors. The advantage of using non-differenced data is more pronounced when the forecasting horizon is long and the number of factors is large. The forecasting method based on differenced data can perform better only if the following conditions hold simultaneously: (i) a researcher can employ many $I(0)$ variables for estimating factor spaces, (ii) the number of factors is large, and (iii) the forecasting horizon is short. Last, we have applied the two competing forecasting methods, one using differenced data and the other using non-differenced ones, to CPI-based U.S. inflation rate and found that forecasts using non-differenced data outperform those using differenced data except for one-month forecasting horizon. For most cases, the Diebold-Mariano test statistics indicate that the performance difference of the two forecasts is statistically significant.

A practical implication of this paper is that using non-differenced data is advisable for the factor-based forecasting especially for long-run forecasting. This paper contains empirical results only for the U.S. inflation rate, and it remains to be seen whether using non-differenced data is advantageous for forecasting other important economic variables throughout the world.

There are methods for inflation forecasting that deserve serious considerations. Stock and Watson (2008b) successfully employ the unobserved component stochastic volatility model for inflation forecasting, and Dotsey, Fujita and Stark (2011) confirm that the threshold model is useful. Inoue and Kilian (2008) show that bagging (Breiman, 1996) produces good forecasting results. Wright (2009) demonstrates the usefulness of the Bayesian model-averaging method for inflation forecasting. Tu and Lee (2011) show how to use supervised factor models effectively to forecast inflation. It is beyond the scope of this paper to compare these methods with the one proposed here, but it would be a meaningful exercise to find which methods provide the best results in forecasting $I(1)$ inflation.

Appendix I: Proofs of main results

Proof of Lemma 2: (i) Using equation (7), we have

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

If $\frac{T}{N_1} \to 0$ and $\frac{N_1}{T^3} \to 0$ or if $\liminf \frac{N_1}{T^3} \geq \delta > 0$, the second term is $o_p(1)$ as shown in Lemma A.II.1 (i). Moreover, it is straightforward to show as in Lemma B.7 of Choi (2017):

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

and

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

where $\Psi_L = 1 \oplus \left(\left(V^L\right)^{-1/2}\Upsilon^{L\prime}\Sigma_\Lambda^{1/2}\right) \oplus I_K$. 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.

(ii) The OLS estimator from equation (9) gives

$$
\begin{aligned}
\sqrt{T}\left(\hat{\delta}^D - \delta^D\right) &= \left(T^{-1}\sum_{t=1}^{T-h}\tilde{P}_t\tilde{P}_t'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\tilde{P}_t\Delta\epsilon_{t+h} \\
&\quad + \left(T^{-1}\sum_{t=1}^{T-h}\tilde{P}_t\tilde{P}_t'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\tilde{P}_t\left(H_D'u_t - \tilde{u}_t\right)' H_D^{-1}\alpha \\
&= \left(T^{-1}\sum_{t=1}^{T-h}\tilde{P}_t\tilde{P}_t'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\tilde{P}_t\Delta\epsilon_{t+h} + o_p(1),
\end{aligned}
$$

where Lemma A.II.1 (ii) is used for the second equality. In addition, as in Bai and Ng (2006), we obtain under Assumption 6

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T-h}\tilde{P}_t\tilde{P}_t' &= T^{-1}\left(1\oplus H_D'\oplus I_K\right)\sum_{t=1}^{T-h}\Delta L_t\Delta L_t'\left(1\oplus H_D\oplus I_K\right) + o_p(1) \\
&\xrightarrow{p} \Psi_L\Sigma_{\Delta L}\Psi_L'
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\tilde{P}_t\Delta\epsilon_{t+h} &= \frac{1}{\sqrt{T}}\left(1\oplus H_D'\oplus I_K\right)\sum_{t=1}^{T-h}\Delta L_t\Delta\epsilon_{t+h} + o_p(1) \\
&\xrightarrow{d} \Psi_D N(0,\Sigma_{\Delta\epsilon L}),
\end{aligned}
$$

from which the stated result follows. ■

Proof of Theorem 3: (i) From equation (10), we obtain

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

Because $H_L = \left(\frac{\Lambda'\Lambda}{N_1}\right)\left(\frac{F'\tilde{F}}{T^2}\right)\left(V^L_{N_1T}\right)^{-1} \xrightarrow{d} \Sigma_\Lambda Q^{L\prime}V^{L-1} = \Sigma_\Lambda^{1/2}\Upsilon^L\left(V^L\right)^{-1/2}$, Lemma 2 yields $\left(1\oplus H_L'\oplus I_K\right)D_T\left(\hat{\delta}^L - \delta^L\right) \xrightarrow{d} \Sigma_L^{-1}\Sigma_{\epsilon L}^{1/2}N(0,I)$. Moreover, Lemma 1 implies $\tilde{L}_T = \begin{bmatrix} 1 \\ H_L'F_T + O_p\left(\frac{1}{\min\{\sqrt{N_1},T^{3/2}\}}\right) \\ Z_T \end{bmatrix}$, which gives

$$
\begin{aligned}
\left(1 \oplus H_L'^{-1} \oplus I_K\right) \tilde{L}_T &= \begin{bmatrix} 1 \\ F_T + O_p\left(\frac{1}{\min\{\sqrt{N_1}, T^{3/2}\}}\right) \\ Z_T \end{bmatrix} \\
&= L_T + o_p(1).
\end{aligned}
$$

Thus, conditional on $\{L_t\}_{t=1}^T$, $\widehat{Var}(A) = L_T' D_T^{-1} \Sigma_L^{-1} \Sigma_{\epsilon L} \Sigma_L^{-1} D_T^{-1} L_T$. Next, consider B. Since $H_L'^{-1}\left(\sqrt{N_1}\left(\tilde{F}_T - H_L' F_T\right)\right) \xrightarrow{d} N\left(0, \Sigma_\Lambda^{-1} \Sigma_T \Sigma_\Lambda^{-1}\right)$ by Lemma 1, $\widehat{Var}(B) = \frac{1}{N} \alpha' \Sigma_\Lambda^{-1} \Sigma_T \Sigma_\Lambda^{-1} \alpha$. Note that the asymptotic distribution of A is driven by $\{L_t, \epsilon_t\}_{t=1}^{T-h}$ while that of B rests on $\{e_t\}_{t=1}^T$. Thus, by Assumption 6 (i), A and B are asymptotically uncorrelated. Moreover, A and ϵ_{T+h} are asymptotically uncorrelated because $E\left(\epsilon_{t+h}\epsilon_{T+h} | L_1, \cdots, L_{T-h}\right) = 0$ for $t = 1, \cdots, T-h$. B and ϵ_{T+h} are asymptotically uncorrelated due to Assumption 6 (i). Thus, the stated result follows.

(ii) This can be proven using the same arguments as for part (i). ■

Appendix II: Auxiliary Lemmas

Lemma A.II.1 *Suppose that Assumptions 1, 2, 3, 5 and 6 hold. Then,*
(i) $D_T^{-1} \sum_{t=1}^{T-h} \tilde{L}_t \left(H_L' F_t - \tilde{F}_t\right)' H_L^{-1} \alpha = O_p\left(\frac{\sqrt{T}}{\min\{\sqrt{N_1}, T^{3/2}\}}\right) + O_p\left(\frac{T}{\min\{N_1, T^3\}}\right)$.
(ii) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \tilde{P}_t \left(H_D' u_t - \tilde{u}_t\right)' H_D^{-1} \alpha = O_p\left(\frac{\sqrt{T}}{\min\{\sqrt{N}, T\}}\right) + O_p\left(\frac{T}{\min\{N, T^2\}}\right)$.

Proof: (i) Let $M_t = \begin{pmatrix} 1 \\ H_L' F_t \\ Z_t \end{pmatrix}$ and $D_T = \sqrt{T} \oplus diag(T, ..., T) \oplus E_T$, where E_T is a diagonal matrix of dimension K. Write

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

Because $H_L' F_t - \tilde{F}_t = O_p(\frac{1}{\min\{\sqrt{N_1}, T^{3/2}\}})$ as in Lemma 1,

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

Let $A_2 = \begin{pmatrix} A_{21} \\ A_{22} \\ A_{23} \end{pmatrix} \begin{matrix} 1 \\ r \\ K \end{matrix}$. Because $\frac{1}{\sqrt{T}} F_t = O_p(1)$, we have for any t

$$A_{21} = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left(H_L' F_t - \tilde{F}_t\right)'\right) H_L^{-1}\alpha = O_p\left(\frac{\sqrt{T}}{\min\{\sqrt{N_1}, T^{3/2}\}}\right) \tag{A.II.2}$$

$$A_{22} = H_L'\left(\frac{1}{T}\sum_{t=1}^{T-h} F_t\left(H_L' F_t - \tilde{F}_t\right)'\right) H_L^{-1}\alpha = O_p\left(\frac{\sqrt{T}}{\min\{\sqrt{N_1}, T^{3/2}\}}\right). \tag{A.II.3}$$

In addition, the Cauchy-Schwarz inequality gives

$$\begin{aligned} \|A_{23}\| &= \left\|E_T^{-1}\sum_{t=1}^{T-h} Z_t\left(\tilde{F}_t - H_L' F_t\right)'\right\| \\ &\leq \left\|E_T^{-1}\sum_{t=1}^{T-h} Z_t Z_t' E_T^{-1}\right\|\left\|\sum_{t=1}^{T-h}\left(\tilde{F}_t - H_L' F_t\right)\left(\tilde{F}_t - H_L' F_t\right)'\right\| \\ &= O_p\left(\frac{T}{\min\{N_1, T^3\}}\right). \end{aligned} \tag{A.II.4}$$

The stated result follows from relations (A.II.1), (A.II.2), (A.II.3) and (A.II.4).

(ii) Use the same method as for part (i). ■

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.

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 (2002). Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

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.

Breiman, L. (1996). Bagging predictors, Machine Learning 24, 123–140.

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.

Bühlmann, P. and S. van de Geer (2011). Statistics for high-dimensional data: methods, theory and applications, Springer-Verlag.

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

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

Diebold, F.X. (2016). Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold–Mariano tests, Journal of Business and Economic Statistics, 33, 1-9.

Diebold, F.X. and R.S. Mariano (2002). Comparing predictive accuracy, Journal of Business and Economic Statistics, 20, 134-144.

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.

Harvey, D.I., S.J. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13, 281-291.

Hossfeld, O. (2010). US Money demand, monetary overhang, and inflation, INFER Working Paper 2010.4.

Inoue A. and L. Kilian (2008). How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. Journal of the American Statistical Association, 103, 511–522.

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

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.N. Durlauf. (1986). Multiple time series regression with integrated processes, Review of Economic Studies, 53, 473–495.

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, Vol. 41, No. 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 (2008a). 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 (2008b). Phillips curve inflation forecasts. No 14322, NBER Working Papers, National Bureau of Economic Research.

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.

Tu, Y and T. Lee (2011). Forecasting using supervised factor models. Working Paper, University of California, Riverside.

Wright, J.H. (2009). Forecasting US inflation by Bayesian model averaging. Journal of Forecasting, 28, 131–144.

Table 1: Root mean squared forecasting errors

$(T,N)=(50,80)$

N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D	N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D	N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D
0.1	1	1	1.0179	1.2549	0.8111	0.1	1	3	1.0697	1.2745	0.8393	0.1	1	5	1.3532	1.3595	0.9954
0.33	1	1	1.0001	1.2835	0.7792	0.33	1	3	1.0473	1.3364	0.7837	0.33	1	5	1.2960	1.4106	0.9188
0.5	1	1	0.9646	1.2744	0.7569	0.5	1	3	1.0422	1.3321	0.7824	0.5	1	5	1.1554	1.4531	0.7951
0.66	1	1	0.9762	1.2361	0.7898	0.66	1	3	1.0067	1.3321	0.7557	0.66	1	5	1.0620	1.3432	0.7906
0.9	1	1	0.9510	1.2657	0.7513	0.9	1	3	0.9698	1.3130	0.7386	0.9	1	5	1.0251	1.3601	0.7537
0.1	3	1	1.0074	1.4233	0.7078	0.1	3	3	1.1085	1.5369	0.7212	0.1	3	5	1.4104	1.9218	0.7339
0.33	3	1	0.9474	1.3925	0.6803	0.33	3	3	1.0452	1.5508	0.6739	0.33	3	5	1.2802	1.8494	0.6922
0.5	3	1	0.9726	1.4009	0.6943	0.5	3	3	1.0121	1.5683	0.6453	0.5	3	5	1.1000	1.8465	0.5957
0.66	3	1	0.9743	1.4004	0.6958	0.66	3	3	0.9848	1.5499	0.6354	0.66	3	5	1.0415	1.6923	0.6155
0.9	3	1	0.9748	1.3876	0.7025	0.9	3	3	0.9473	1.4512	0.6528	0.9	3	5	1.0268	1.6362	0.6275
0.1	6	1	0.9827	1.4174	0.6933	0.1	6	3	1.0830	1.8885	0.5735	0.1	6	5	1.3632	2.6139	0.5215
0.33	6	1	0.9672	1.4197	0.6813	0.33	6	3	1.0083	1.7997	0.5603	0.33	6	5	1.2582	2.4243	0.5190
0.5	6	1	0.9814	1.4122	0.6950	0.5	6	3	1.0145	1.8343	0.5531	0.5	6	5	1.1238	2.3556	0.4771
0.66	6	1	0.9722	1.4045	0.6922	0.66	6	3	0.9683	1.6892	0.5732	0.66	6	5	1.0248	2.0243	0.5062
0.9	6	1	0.9578	1.4433	0.6636	0.9	6	3	0.9518	1.5644	0.6084	0.9	6	5	1.0002	1.8322	0.5459
0.1	12	1	0.9720	1.4965	0.6496	0.1	12	3	1.0310	2.2585	0.4565	0.1	12	5	1.2890	3.2782	0.3932
0.33	12	1	0.9617	1.5173	0.6338	0.33	12	3	1.0069	2.1586	0.4665	0.33	12	5	1.2208	2.9368	0.4157
0.5	12	1	0.9626	1.5015	0.6410	0.5	12	3	0.9926	2.1504	0.4616	0.5	12	5	1.0678	3.0275	0.3527
0.66	12	1	0.9237	1.4563	0.6343	0.66	12	3	0.9696	1.9156	0.5062	0.66	12	5	0.9935	2.4800	0.4006
0.9	12	1	0.9513	1.4293	0.6656	0.9	12	3	0.9315	1.7860	0.5215	0.9	12	5	0.9807	2.0884	0.4696

$(T,N)=(50,150)$

N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D	N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D	N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D
0.1	1	1	1.0054	1.2714	0.7908	0.1	1	3	1.1076	1.2941	0.8559	0.1	1	5	1.6695	1.3042	1.2800
0.33	1	1	0.9504	1.2480	0.7615	0.33	1	3	0.9888	1.2920	0.7653	0.33	1	5	1.1373	1.3434	0.8466
0.5	1	1	0.9810	1.2855	0.7631	0.5	1	3	1.0084	1.3412	0.7518	0.5	1	5	1.0380	1.3440	0.7723
0.66	1	1	0.9571	1.2431	0.7699	0.66	1	3	0.9562	1.3152	0.7271	0.66	1	5	1.0043	1.3021	0.7713
0.9	1	1	0.9976	1.3045	0.7648	0.9	1	3	0.9598	1.3336	0.7197	0.9	1	5	0.9667	1.3543	0.7139
0.1	3	1	0.9969	1.4027	0.7107	0.1	3	3	1.1177	1.5292	0.7309	0.1	3	5	1.7415	1.8668	0.9329
0.33	3	1	0.9693	1.4006	0.6921	0.33	3	3	1.0098	1.5055	0.6708	0.33	3	5	1.1350	1.7248	0.6580
0.5	3	1	0.9632	1.3829	0.6965	0.5	3	3	0.9561	1.5137	0.6316	0.5	3	5	1.0163	1.6130	0.6300
0.66	3	1	0.9478	1.3540	0.6999	0.66	3	3	0.9563	1.4635	0.6534	0.66	3	5	1.0096	1.6241	0.6216
0.9	3	1	0.9703	1.3961	0.6950	0.9	3	3	0.9387	1.4385	0.6525	0.9	3	5	0.9780	1.5993	0.6115
0.1	6	1	0.9699	1.3997	0.6929	0.1	6	3	1.1239	1.8395	0.6110	0.1	6	5	1.7536	2.5276	0.6938
0.33	6	1	0.9470	1.4420	0.6567	0.33	6	3	0.9449	1.6506	0.5725	0.33	6	5	1.1282	2.2481	0.5018
0.5	6	1	0.9699	1.4283	0.6791	0.5	6	3	0.9658	1.7250	0.5599	0.5	6	5	0.9992	1.9578	0.5104
0.66	6	1	0.9200	1.4020	0.6562	0.66	6	3	0.9445	1.6237	0.5817	0.66	6	5	1.0051	1.9660	0.5112
0.9	6	1	0.9453	1.3837	0.6831	0.9	6	3	0.9267	1.5800	0.5865	0.9	6	5	0.9523	1.7838	0.5339
0.1	12	1	0.9275	1.4596	0.6355	0.1	12	3	1.0806	2.2367	0.4831	0.1	12	5	1.6587	3.2515	0.5101
0.33	12	1	0.9411	1.4841	0.6341	0.33	12	3	0.9225	1.9745	0.4672	0.33	12	5	1.0967	2.7660	0.3965
0.5	12	1	0.9685	1.4383	0.6733	0.5	12	3	0.9811	1.9074	0.5144	0.5	12	5	0.9829	2.3795	0.4131
0.66	12	1	0.9125	1.4850	0.6145	0.66	12	3	0.9105	1.7828	0.5107	0.66	12	5	0.9529	2.2439	0.4247
0.9	12	1	0.9300	1.4643	0.6351	0.9	12	3	0.9021	1.7308	0.5212	0.9	12	5	0.9210	2.0235	0.4552

$(T, N) = (150, 80)$

N_1 / N	h	r	$RMSE_h^L$	$RMSE_h^D$	L / D	N_1 / N	h	r	$RMSE_h^L$	$RMSE_h^D$	L / D	N_1 / N	h	r	$RMSE_h^L$	$RMSE_h^D$	L / D
0.1	1	1	1.0129	1.2776	0.7928	0.1	1	3	1.1636	1.3587	0.8564	0.1	1	5	1.5144	1.4117	1.0727
0.33	1	1	0.9996	1.2738	0.7847	0.33	1	3	1.1069	1.3169	0.8405	0.33	1	5	1.3684	1.3928	0.9825
0.5	1	1	0.9789	1.2597	0.7771	0.5	1	3	1.0877	1.3702	0.7938	0.5	1	5	1.2302	1.4122	0.8712
0.66	1	1	1.0042	1.3133	0.7646	0.66	1	3	1.0094	1.3154	0.7674	0.66	1	5	1.1636	1.4581	0.7980
0.9	1	1	1.0062	1.2887	0.7808	0.9	1	3	1.0120	1.3484	0.7505	0.9	1	5	1.1025	1.4278	0.7721
0.1	3	1	1.0430	1.4512	0.7187	0.1	3	3	1.1496	1.5739	0.7304	0.1	3	5	1.5406	1.8165	0.8481
0.33	3	1	0.9793	1.4128	0.6932	0.33	3	3	1.0991	1.5663	0.7017	0.33	3	5	1.4074	1.7698	0.7952
0.5	3	1	0.9721	1.3994	0.6947	0.5	3	3	1.0906	1.5451	0.7058	0.5	3	5	1.2505	1.8110	0.6905
0.66	3	1	1.0072	1.4337	0.7025	0.66	3	3	1.0444	1.5018	0.6955	0.66	3	5	1.1785	1.6966	0.6946
0.9	3	1	0.9832	1.4072	0.6987	0.9	3	3	1.0078	1.4574	0.6915	0.9	3	5	1.1261	1.6498	0.6825
0.1	6	1	1.0300	1.4515	0.7096	0.1	6	3	1.1546	1.8193	0.6346	0.1	6	5	1.5651	2.3756	0.6588
0.33	6	1	1.0102	1.4540	0.6948	0.33	6	3	1.1073	1.7853	0.6202	0.33	6	5	1.4412	2.3305	0.6184
0.5	6	1	1.0162	1.4137	0.7189	0.5	6	3	1.1096	1.8166	0.6108	0.5	6	5	1.2819	2.3110	0.5547
0.66	6	1	0.9906	1.4278	0.6938	0.66	6	3	1.0718	1.6762	0.6394	0.66	6	5	1.1921	2.0185	0.5906
0.9	6	1	1.0386	1.4783	0.7026	0.9	6	3	1.0322	1.5726	0.6564	0.9	6	5	1.1481	1.8308	0.6271
0.1	12	1	1.0259	1.4973	0.6852	0.1	12	3	1.1481	2.2460	0.5112	0.1	12	5	1.6145	3.2782	0.4925
0.33	12	1	0.9676	1.4700	0.6582	0.33	12	3	1.0954	2.2107	0.4955	0.33	12	5	1.4297	3.1370	0.4558
0.5	12	1	1.0154	1.4594	0.6958	0.5	12	3	1.0869	2.2346	0.4864	0.5	12	5	1.2470	3.0620	0.4073
0.66	12	1	0.9975	1.4280	0.6985	0.66	12	3	1.0372	1.8180	0.5705	0.66	12	5	1.1641	2.5444	0.4575
0.9	12	1	1.0259	1.4624	0.7015	0.9	12	3	1.0382	1.6838	0.6166	0.9	12	5	1.0854	1.9695	0.5511

$(T,N)=(150,150)$

N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D	N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D	N_1/N	h	r	$RMSE_h^L$	$RMSE_h^D$	L/D
0.1	1	1	0.9849	1.2483	0.7890	0.1	1	3	1.1947	1.3429	0.8896	0.1	1	5	1.7384	1.3827	1.2572
0.33	1	1	1.0085	1.2985	0.7767	0.33	1	3	1.0501	1.3369	0.7854	0.33	1	5	1.2359	1.4134	0.8744
0.5	1	1	0.9973	1.3076	0.7627	0.5	1	3	1.0175	1.3526	0.7523	0.5	1	5	1.1238	1.3954	0.8053
0.66	1	1	1.0052	1.2794	0.7857	0.66	1	3	0.9941	1.3460	0.7386	0.66	1	5	1.0987	1.3983	0.7857
0.9	1	1	0.9820	1.2892	0.7618	0.9	1	3	1.0138	1.3657	0.7424	0.9	1	5	1.0826	1.4245	0.7600
0.1	3	1	0.9870	1.3853	0.7125	0.1	3	3	1.2041	1.5273	0.7884	0.1	3	5	2.0050	1.8072	1.1094
0.33	3	1	0.9990	1.3799	0.7240	0.33	3	3	1.0243	1.4699	0.6968	0.33	3	5	1.2833	1.7051	0.7527
0.5	3	1	0.9988	1.4156	0.7056	0.5	3	3	1.0336	1.4927	0.6925	0.5	3	5	1.1391	1.5890	0.7169
0.66	3	1	0.9916	1.4265	0.6951	0.66	3	3	1.0016	1.4816	0.6760	0.66	3	5	1.1035	1.5972	0.6909
0.9	3	1	0.9993	1.4386	0.6946	0.9	3	3	1.0241	1.4688	0.6973	0.9	3	5	1.0529	1.5339	0.6864
0.1	6	1	1.0122	1.4531	0.6966	0.1	6	3	1.1959	1.7005	0.7033	0.1	6	5	2.0842	2.2408	0.9301
0.33	6	1	1.0112	1.4489	0.6979	0.33	6	3	1.0167	1.6608	0.6122	0.33	6	5	1.2932	2.1332	0.6062
0.5	6	1	0.9891	1.4417	0.6861	0.5	6	3	1.0489	1.6235	0.6461	0.5	6	5	1.1004	1.8583	0.5922
0.66	6	1	1.0192	1.4445	0.7056	0.66	6	3	1.0179	1.5840	0.6426	0.66	6	5	1.0953	1.8235	0.6007
0.9	6	1	0.9992	1.4266	0.7004	0.9	6	3	0.9993	1.4947	0.6685	0.9	6	5	1.0739	1.7534	0.6125
0.1	12	1	0.9601	1.4206	0.6758	0.1	12	3	1.1680	2.1424	0.5452	0.1	12	5	2.0664	3.1348	0.6592
0.33	12	1	0.9964	1.4462	0.6889	0.33	12	3	1.0140	1.8805	0.5392	0.33	12	5	1.2326	2.8552	0.4317
0.5	12	1	0.9996	1.4515	0.6887	0.5	12	3	1.0096	1.8655	0.5412	0.5	12	5	1.0824	2.3776	0.4553
0.66	12	1	0.9698	1.4199	0.6830	0.66	12	3	0.9998	1.7190	0.5816	0.66	12	5	1.1011	2.2568	0.4879
0.9	12	1	0.9778	1.4208	0.6882	0.9	12	3	0.9821	1.6690	0.5884	0.9	12	5	1.0732	1.9738	0.5437

Table 2: Performance of forecasts

Note: 1. Numbers in parentheses are p-values. 2. The unemployment and the term spread are denoted by ur. and sprd., respectively.

$h=1$

Model	DM ($L_1(\xi_t)$)	DM ($L_2(\xi_t)$)	$HLN-DM$ ($L_1(\xi_t)$)	$HLN-DM$ ($L_2(\xi_t)$)	$RMSE_h^L$	$RMSE_h^D$	$\frac{RMSE_h^L}{RMSE_h^D}$	MAE_h^L	MAE_h^D	$\frac{MAE_h^L}{MAE_h^D}$
(i) none	-2.6293	-3.3644	-2.6251	-3.3590	2.4731	**2.2306**	1.1087	1.8494	**1.6182**	1.1429
	(0.0086)	(0.0008)	(0.0090)	(0.0009)						
(ii) ur.	-2.4415	-2.9930	-2.4375	-2.9882	2.4482	**2.2132**	1.1062	1.8246	**1.6111**	1.1325
	(0.0146)	(0.0028)	(0.0152)	(0.0030)						
(iii) sprd.	-2.6904	-3.3894	-2.6861	-3.3840	2.4697	**2.2198**	1.1125	1.8487	**1.6102**	1.1481
	(0.0071)	(0.0007)	(0.0075)	(0.0008)						
(iv) ur.+sprd.	-2.5072	-3.1385	-2.5031	-3.1334	2.4540	**2.2086**	1.1111	1.8296	**1.6044**	1.1403
	(0.0122)	(0.0017)	(0.0127)	(0.0019)						

$h=3$

Model	DM ($L_1(\xi_t)$)	DM ($L_2(\xi_t)$)	$HLN-DM$ ($L_1(\xi_t)$)	$HLN-DM$ ($L_2(\xi_t)$)	$RMSE_h^L$	$RMSE_h^D$	$\frac{RMSE_h^L}{RMSE_h^D}$	MAE_h^L	MAE_h^D	$\frac{MAE_h^L}{MAE_h^D}$
(i) none	2.6651	2.5563	2.6436	2.5356	**2.6511**	3.1948	0.8298	**1.9576**	2.2895	0.8551
	(0.0077)	(0.0106)	(0.0081)	(0.0111)						
(ii) ur.	2.6652	2.7846	2.6437	2.7621	**2.6273**	3.1678	0.8294	**1.9270**	2.2807	0.8449
	(0.0077)	(0.0054)	(0.0081)	(0.0057)						
(iii) sprd.	2.6030	2.6223	2.5820	2.6011	**2.6143**	3.1565	0.8282	**1.9169**	2.2582	0.8489
	(0.0092)	(0.0087)	(0.0097)	(0.0092)						
(iv) ur.+sprd.	2.6963	2.7526	2.6746	2.7304	**2.5965**	3.1416	0.8265	**1.9137**	2.2591	0.8471
	(0.0070)	(0.0059)	(0.0074)	(0.0063)						

$h=6$

Model	DM ($L_1(\xi_t)$)	DM ($L_2(\xi_t)$)	$HLN-DM$ ($L_1(\xi_t)$)	$HLN-DM$ ($L_2(\xi_t)$)	$RMSE_h^L$	$RMSE_h^D$	$\frac{RMSE_h^L}{RMSE_h^D}$	MAE_h^L	MAE_h^D	$\frac{MAE_h^L}{MAE_h^D}$
(i) none	3.5380	3.1562	3.4746	3.0996	**2.6630**	3.3609	0.7924	**2.0016**	2.4725	0.8096
	(0.0004)	(0.0016)	(0.0005)	(0.0018)						
(ii) ur.	4.3606	4.8081	4.2824	4.7220	**2.5604**	3.2868	0.7790	**1.8613**	2.4209	0.7689
	(0.0000)	(0.0000)	(0.0000)	(0.0000)						
(iii) sprd.	3.4708	3.3878	3.4086	3.3271	**2.6035**	3.3097	0.7866	**1.9257**	2.4359	0.7905
	(0.0005)	(0.0007)	(0.0006)	(0.0008)						
(iv) ur.+sprd.	3.6268	4.1947	3.5618	4.1195	**2.5487**	3.2260	0.7900	**1.8461**	2.3971	0.7701
	(0.0003)	(0.0000)	(0.0003)	(0.0000)						

$h=12$

Model	DM ($L_1(\xi_t)$)	DM ($L_2(\xi_t)$)	$HLN-DM$ ($L_1(\xi_t)$)	$HLN-DM$ ($L_2(\xi_t)$)	$RMSE_h^L$	$RMSE_h^D$	$\frac{RMSE_h^L}{RMSE_h^D}$	MAE_h^L	MAE_h^D	$\frac{MAE_h^L}{MAE_h^D}$
(i) none	1.5446	1.2841	1.4856	1.2350	**3.0938**	3.6440	0.8490	**2.3734**	2.7485	0.8635
	(0.1224)	(0.1991)	(0.1235)	(0.2001)						
(ii) ur.	2.0551	1.9991	1.9766	1.9228	**2.8976**	3.6264	0.7990	**2.1943**	2.7440	0.7997
	(0.0399)	(0.0456)	(0.0407)	(0.0465)						
(iii) sprd.	1.8365	1.6125	1.7663	1.5509	**3.0743**	3.6343	0.8459	**2.3353**	2.7208	0.8583
	(0.0663)	(0.1068)	(0.0673)	(0.1079)						
(iv) ur.+sprd.	2.7257	2.8582	2.6216	2.7490	**2.8901**	3.4590	0.8355	**2.1810**	2.6303	0.8292
	(0.0064)	(0.0043)	(0.0068)	(0.0046)						

$h = 24$

Model	DM ($L_1(\xi_t)$)	DM ($L_2(\xi_t)$)	$HLN-DM$ ($L_1(\xi_t)$)	$HLN-DM$ ($L_2(\xi_t)$)	$RMSE_h^L$	$RMSE_h^D$	$\frac{RMSE_h^L}{RMSE_h^D}$	MAE_h^L	MAE_h^D	$\frac{MAE_h^L}{MAE_h^D}$
(i) none	2.2242	1.7085	2.0433	1.5696	**3.2029**	4.1276	0.7760	**2.6210**	3.1757	0.8253
	(0.0261)	(0.0875)	(0.0269)	(0.0886)						
(ii) ur.	2.2052	1.8390	2.0259	1.6894	**3.0321**	3.9321	0.7711	**2.4238**	3.0621	0.7915
	(0.0274)	(0.0659)	(0.0282)	(0.0670)						
(iii) sprd.	2.1913	2.1139	2.0131	1.9420	**3.2803**	4.1654	0.7875	**2.6609**	3.2916	0.8084
	(0.0284)	(0.0345)	(0.0292)	(0.0354)						
(iv) ur.+sprd.	2.1304	2.1408	1.9572	1.9667	**3.0939**	4.0135	0.7709	**2.4679**	3.2168	0.7672
	(0.0331)	(0.0323)	(0.0340)	(0.0331)						

Figure 1: CPI-based U.S. inflation rate

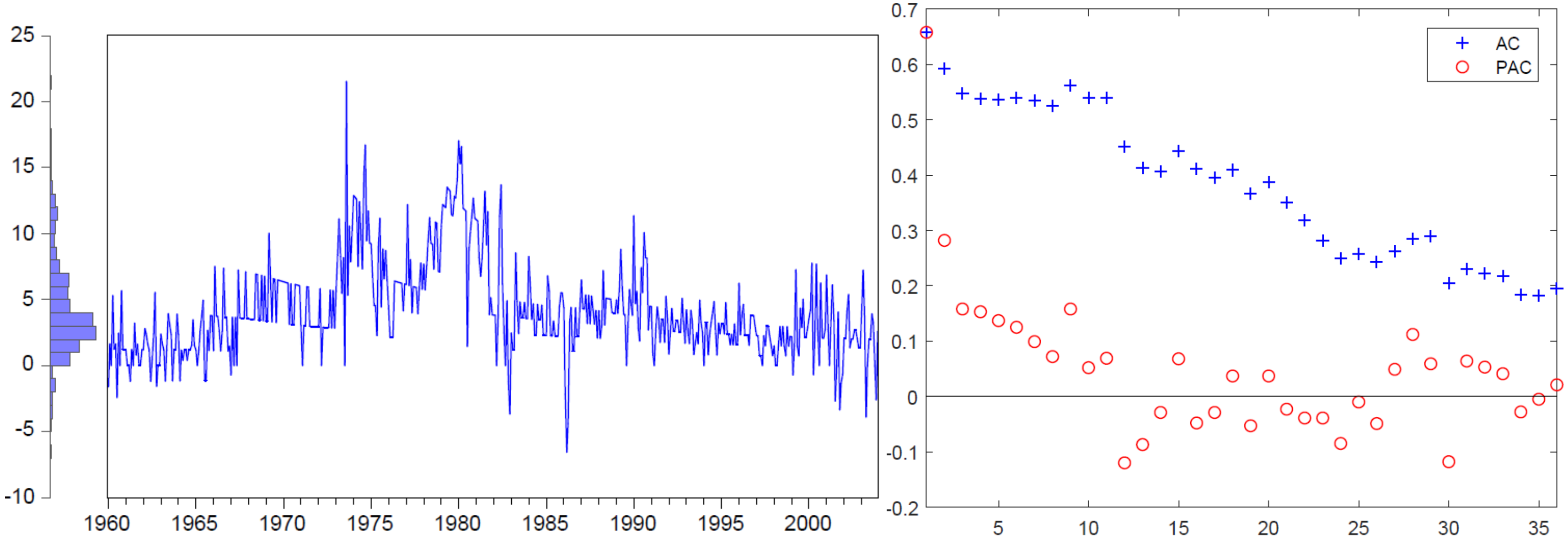
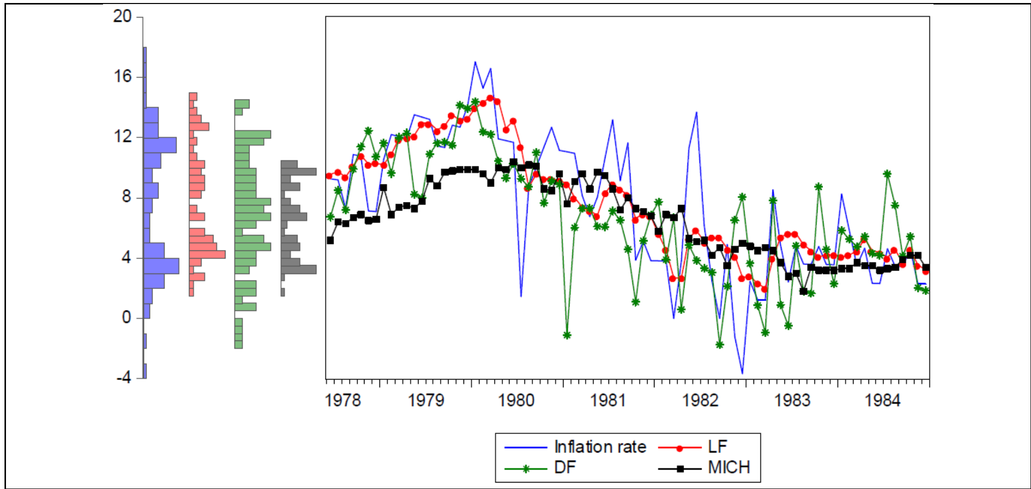
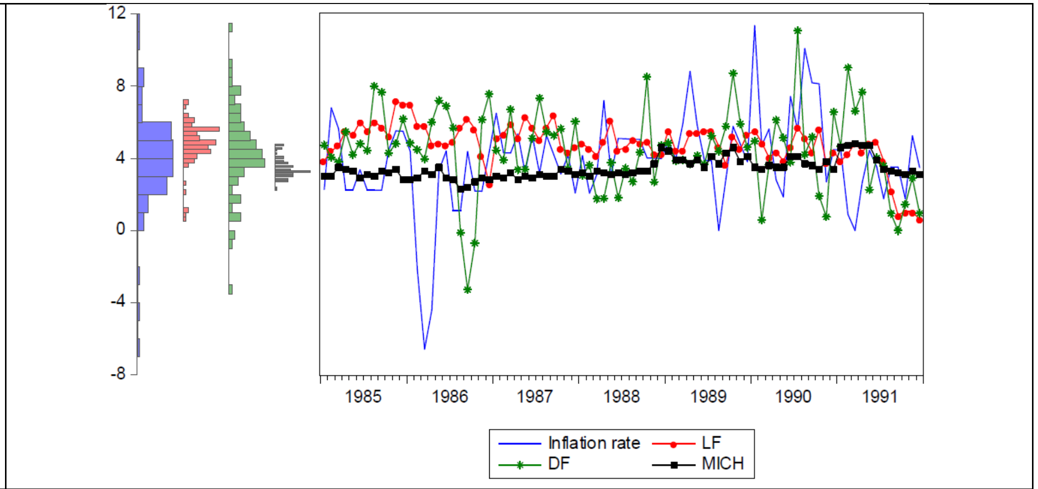


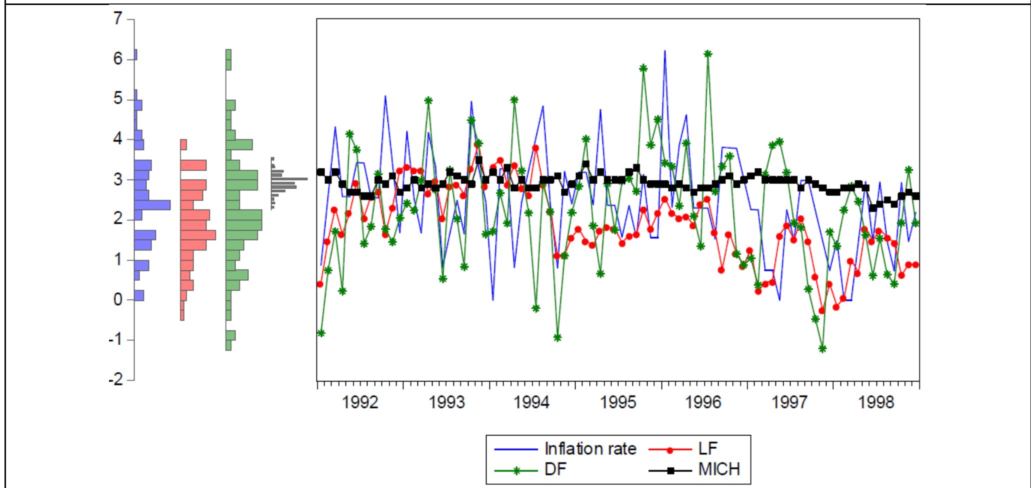
Figure 2: Predicted inflation rates



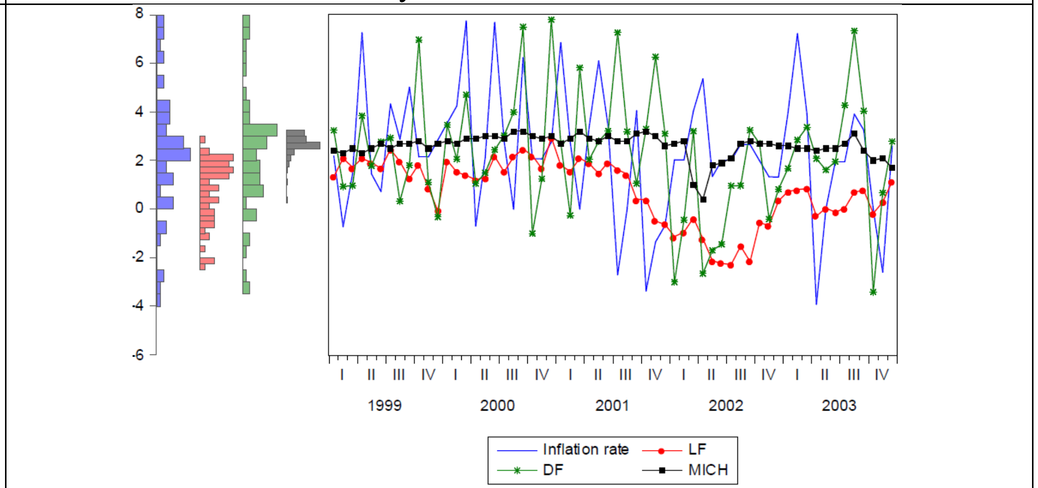
June 1978 ~ December 1984



January 1985 ~ December 1991



January 1992 ~ December 1998



January 1999 ~ December 2003