Forecasting Output Growth and Inflation: How to Use Information in the Yield Curve*

Huiyu Huang[†]
Department of Economics
University of California, Riverside

Tae-Hwy Lee[‡]
Department of Economics
University of California, Riverside

Canlin Li[§]
Anderson Graduate School of Management
University of California, Riverside

December 2006

Abstract

We examine how one can use information in the entire yield curve to improve forecasts of output growth and inflation. First, we consider two ways of forming the forecast model: combining forecasts (CF) with each individual forecast obtained from using one yield at a time and combining information (CI) from the entire yields into one super model. Next, for both the CF and CI methods, we consider two alternative factorizing frameworks: principal component (PC) method and Nelson-Siegel (NS) exponential components modeling of the yield curve. The main contribution of this paper is to introduce the NS factorizing framework for CF (which we term as CF-NS), that is a *new* way of combining forecasts when the yield curve is used for forecasting. In out-of-sample forecasting of monthly output growth and inflation, we find that the CF method is more successful than the CI method, CF-NS is more stable than CF using PC factors, and that CF-NS using two or three factors (capturing the level, slope and curvature of the yield curve) generally works the best in out-of-sample forecasting of output growth and inflation.

Key Words: Combining forecasts; Level, slope, and curvature of the yield curve; Nelson-Siegel factors; Pairs trading; Partial least squares; Principal components.

JEL Classification: C5, E4, G1.

---

*We would like to thank Gloria González-Rivera, Clive Granger, Jim Hamilton, Jinyong Hahn, Cheng Hsiao, Mike McCracken, Roger Moon, Eddie Qian, Aaron Smith, Jim Stock, Allan Timmermann, Yixiao Sun, Aman Ullah, Hal White, as well as seminar participants at Federal Reserve Bank of St. Louis, North Carolina State University, PanAgora Asset Management, University of Southern California, University of California, Riverside, and University of California, San Diego for helpful discussions and comments. All errors are our own.

[†]Department of Economics, University of California, Riverside, CA 92521-0427, E-mail: huiyu.huang@email.ucr.edu.

[‡]Corresponding author. Department of Economics, University of California, Riverside, CA 92521-0427, E-mail: taelee@ucr.edu, phone: (951) 827-1509.

[§]Anderson Graduate School of Management, University of California, Riverside, CA 92521-0203, E-mail: canlin.li@ucr.edu, phone: (951) 827-2325.

1 Introduction

The usefulness of yield curve on predicting macroeconomic activities has been long documented in the literature with many different points of the yield curve and various methodologies examined. For example Stock and Watson (1989) find that two interest rate spreads, the difference between the six month commercial paper rate and six-month Treasury bill rate, and the difference between the ten year and one year Treasury bond rates, are good predictor of real activities, thus contributing to their Index of Leading Indicators. Bernanke (1990), Friedman and Kuttner (1991), Estrella and Hardouvelis (1991), and Kozicki (1997), among many others, have then investigated a variety of yields and yield spreads individually on their ability to forecasting macro variables. Hamilton and Kim (2002) and Diebold, Piazzesi, and Rudebusch (2005) provide a brief summary on this line of research and the link between the yield curve and the macroeconomic activities.

Recently, researchers start to study the entire yield curve for its predictive power of real activity and inflation. In a sequence of papers, Stock and Watson (1999, 2002, 2004, 2005a) investigate forecasts of output (real GDP or Industrial Production) growth and/or inflation using over a hundred of economic indicators, including ten interest rates and nine yield spreads, by various methods. Ang, Piazzesi and Wei (2006) suggest the use of the short rate, the five year to three month yield spread, and lagged GDP growth in forecasting GDP growth out-of-sample. The choice of these two yield curve characteristics, as they argue, is because they have almost one-to-one correspondence with the first two principal components of short rate and five yield spreads that account for 99.7% of quarterly yield curve variation. Another group of studies focus on modeling and/or forecasting yield curve itself. Diebold and Li (2006), following Nelson and Siegel (1987), use a modified three-factor model to capture the dynamics of the entire yield curve and apply to yield curve forecasting. They show that the three factors may be interpreted as level, slope and curvature, and find encouraging results on term-structure forecasts at long horizons. Diebold, Rudebusch, and Aruoba (2006) examine the correlations between Nelson-Siegel yield factors and macroeconomic variables. They find that the level factor is highly correlated with inflation, and the slope factor is highly correlated with real activity.

In the mean time, various methodologies on exploring yield information for real activity prediction are proposed, either theory based or non-theory based. Ang and Piazzesi (2003) and Piazzesi (2005) study the role of macroeconomic variables in a no-arbitrage affine yield curve model. Estrella (2005) constructs an analytical rational expectations model to investigate the reasons for the success of the slope of the yield curve (the spread between long-term and short-term government bond rates) on real economic activity and inflation prediction. The model in Ang, Piazzesi and Wei (2006) is a no-arbitrage dynamic model (using lag of GDP growth and yields as regressors) that characterizes expectations of GDP growth. Rudebusch and Wu (2004) provide an example of a macro-finance specification that employs more macroeconomic structure and includes both rational expectations and inertial elements. In contrast, Stock and Watson (1999, 2002, 2004, 2005a), without modeling theoretically the term-structure, advocate methods that aim at solving the large-N problem: forecast combinations and factor models. They compare comprehensively the performance of various types of forecast combinations with factor models, particularly those using principal component (PC) approach, on the large-N predictor information set. They find generally the most accurate forecasts are produced by the factor models. Similar to the factor model approach in its methodologically simple and parsimonious nature, the Nelson-Siegel (NS) exponential components framework of Diebold and Li (2006) involves neither no-arbitrage nor equilibrium approach but shrinkage principle in its essence, whereas being proven to be practically successful in yield curve forecasting.

In this paper, we investigate how one can utilize the entire yield curve information to improve the prediction of macro-economic variables, such as monthly personal income growth and CPI inflation. With parsimony in mind, we consider two alternative factorizing frameworks for incorporating the entire yield curve information. The first factorizing framework is based on the PC method, and the second is based on the NS exponential components of the yield curve.

In Huang and Lee (2006), studies on combination of forecasts (CF, see Timmermann (2005) for a survey) vs. combination of information (CI, forecast generated by combining all the information into one super model) show that the CF method has its merits on out-of-sample forecasting practice. They find that generally CF methods are more successful than CI methods in their empirical application of equity premium prediction, similar to the result of Stock and Watson (2004) in forecasting output growth.

The main contribution of this paper is to combine the advantages of CF with the parsimony of NS framework. To do that, we introduce a new forecasting method (which we term as CF-NS), that applies the NS factorizing framework to CF. The newly proposed CF-NS method is (i) first to combine individual forecasts with three sets of fixed weights that are the three normalized NS exponential loadings corresponding respectively to yield curve level, slope and curvature factors as in Diebold and Li (2006), (ii) then to estimate a regression of the variable to be forecast on these three combined forecasts, and (iii) finally to form the forecast based upon this regression.

See Section 2.2 for details. Out-of-sample forecasting results show that forecasting output growth and inflation generally require two or three NS factors. The CF-NS method using only the first factor (capturing the level of the yield curve) is shown to be the CF method with simple average of forecasts, and it is found to be inferior to the CF-NS with two or three factors (capturing the level, slope and curvature of the yield curve) in out-of-sample forecasting of output growth and inflation. Our empirical study demonstrates that the CF-NS method is generally the best method that uses information in the entire yield curve in forecasting output growth and inflation. The likely reason for its superior performance is due to its ability to capture parsimoniously the slope and curvature information in the yield curve to reflect the uneven contributions from the individual yields in predicting the output growth and inflation. We also discuss that when forecasts are highly correlated and similar, the optimal forecast combination should assign a negative weight to the inferior forecast, which is similar to the "pairs trading" strategy in the finance literature (see, for example, Gatev, Goetzmann, and Rouwenhorst (2006)).

Finally, for inflation forecasting, we compare these yield-curve-based factor forecasting methods with Stock and Watson's (2005b) IMA(1,1) univariate model and find CF with PC or our CF with NS factorizing framework has better performance in long horizons. Yet, we find that it appears to be harder to outperform IMA(1,1) in more recent years, which may indicate diminishing forecasting power of the yield curve within the Great Moderation (periods after mid 1980s, see Kim and Nelson 1999).

The rest of the paper is organized as follows. In Section 2 the two factorizing frameworks (PC and NS) along with the CI and CF schemes are described. Section 3 presents the empirical analysis to examine the out-of-sample forecasting performance of alternative forecasting methods. Section 4 concludes.

2 How to Use Information in the Yield Curve: Methods

In this section, we describe the two forecasting methods, CI and CF, as two alternative ways to combine the entire yield curve information, either directly (CI) or indirectly (CF). In forming the combinations, we focus on two different approaches to factorizing the yield curve: the principal component approach and the Nelson-Siegel framework. The relative advantages of these two competing factorizing approaches will be discussed subsequently in detail. For comparison purpose, we also include forecast combination methods that have been proven to be successful in various empirical applications in the literature (see Stock and Watson 2004, Timmermann 2005), namely the equally weighted CF, median CF, and CF with shrinkage weights.

Let y_{t+h} denote the variable to be forecast (output growth or inflation) using the yield information up to time t, where h denotes the forecast horizon. The predictor vector $\mathbf{x}_t$ contains the information about the entire yield curve at various maturities: $\mathbf{x}_t \equiv (x_t(\tau_1),\ x_t(\tau_2),\ \ldots,\ x_t(\tau_N))'$ where $x_t(\tau_i)$ denotes the yield for maturity of τ_i months at time t $(i = 1, 2, \ldots, N)$. The construction of $\mathbf{x}_t$ for seventeen fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months, is detailed in Section 3.1.

2.1 Factorizing yield curve information directly: CI forecasting schemes

The simplest CI scheme is the OLS using directly $\mathbf{x}_t$ as the regressor set. That is, we run OLS regression $y_{t+h} = (1\ \mathbf{x}_t')\alpha + \varepsilon_{t+h}$ to obtain estimated coefficients $\hat{\alpha}_T$ and the forecast is constructed as $\hat{y}_{T+h} = (1\ \mathbf{x}_T')\hat{\alpha}_T$. Denote this forecasting scheme as "CI-Unrestricted".

CI-PC: When the dimension of $\mathbf{x}_t$ is large, we know that the CI-Unrestricted is very likely of poor out-of-sample performance due to problems such as over-fitting and parameter estimation error. The factor model with Principal Component (PC) approach that factorizes the entire yield curve information is thus promising since it works on mitigating such dimensionality problem through rank reduction. The procedure is as follows:

$$\mathbf{x}_t = \Lambda F_t + v_t, \tag{1}$$

$$y_{t+h} = (1\ F_t')\gamma + u_{t+h}. \tag{2}$$

In equation (1), by applying the classical principal component methodology, the latent common factors $F = (F_1\ F_2\ \cdots\ F_T)'$ is solved by:

$$\hat{F} = X\hat{\Lambda}/N \tag{3}$$

where N is the size of $\mathbf{x}_t$, $X = (\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_T)'$, and factor loading $\hat{\Lambda}$ is set to $\sqrt{N}$ times the eigenvectors corresponding to the r largest eigenvalues of $X'X$ (see, for example, Bai and Ng 2002). Once $\hat{\gamma}_T$ is obtained from (2) by regressing y_t on $(1\ \hat{F}_{t-h}')$ $(t = h+1, h+2, \ldots, T)$, the forecast is constructed as $\hat{y}_{T+h}^{\text{CI-PC}} = (1\ \hat{F}_T')\hat{\gamma}_T$. Denote this forecasting scheme as "CI-PC".

If the true number of factors r is unknown, it could be estimated by minimizing a penalized version of the sum of squared residuals of factor model (1) (divided by NT) according to Bai and Ng (2002). However, Bai and Ng (2002) focus on the true structure of the factor representative part given by equation (1), while the estimation procedure and the asymptotic inference for estimated number of factors have little to do with the forecasting model in equation (2) which however is our main interest. Moreover, to achieve consistency in estimating r, Bai and Ng (2002) require $N \to \infty$ but N in our subsequent empirical study is only seventeen. Therefore, we estimate r by standard information criteria such as AIC and BIC, for which estimated number of factors k is selected by $\min_{1 \leq k \leq k_{max}} IC_k = \ln(SSR(k)/T) + g(T)k$, where k_{max} is the hypothesized upper limit for the true number of factors r (we choose $k_{max} = N = 17$ in our empirical study), $SSR(k)$ is the sum of squared residuals from estimation of the forecasting model (2) using k estimated factors, and the penalty function $g(T) = 2/T$ for AIC and $g(T) = \ln T/T$ for BIC. k can be fixed at some small values like 1, 2, or 3 as well.

CI-NS: Alternatively, to factorize the entire yield curve, one can take the modified Nelson-Siegel (NS) three-factor framework as in Diebold and Li (2006). It proceeds by first fitting the yield curve period by period using the three-factor model:

$$x_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \eta_t(\tau). \tag{4}$$

$(\hat{\beta}_{1t},\ \hat{\beta}_{2t},\ \hat{\beta}_{3t})$ are the three time-varying parameters that are interpreted as factors corresponding to level, slope and curvature, which capture the entire yield curve dynamics over time and are shown to be highly correlated with the empirical level, slope, and curvature of the yield curve (see Diebold and Li (2006) for details).[1] Once they are estimated by running OLS regression of $x_t(\tau)$ on 1, $\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)$, and $\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$, for various maturities τ at each t, we then use them to serve as regressors in the following forecasting regression:

$$y_{t+h} = (1\ \hat{\beta}_{1t}\ \hat{\beta}_{2t}\ \hat{\beta}_{3t})\delta + \epsilon_{t+h}. \tag{5}$$

The CI-NS forecast is computed as: $\hat{y}^{\text{CI-NS}}_{T+h} = (1\ \hat{\beta}_{1T}\ \hat{\beta}_{2T}\ \hat{\beta}_{3T})\hat{\delta}_T$ where $\hat{\delta}_T$ is the estimate of δ using information up to time T. This method is comparable to CI-PC with number of factors fixed at $k = 3$. It differs from CI-PC, however, in that the three NS factors $(\hat{\beta}_{1t},\ \hat{\beta}_{2t},\ \hat{\beta}_{3t})$ bear intuitive interpretations as level, slope and curvature of the yield curve while the first three principal components may not have a clear interpretation. In the empirical section, additionally we consider two alternative CI-NS schemes by including only the level factor $\hat{\beta}_{1t}$ (denoted CI-NS-Level), or only

[1] Similar to CI-NS, alternatively one can use the empirical level, slope, and curvature of yield curve to run a regression of y on these three empirical measures, that are the 10-year yield, the difference between 10-year and 3-month yields, and twice the 2-year yield minus sum of 3-month and 10-year yields, respectively, as defined in Diebold and Li (2006). Empirically we find this method is comparable to the CI-NS method but worse in most cases (thus not reported). Such a result is not surprising by noting that the three empirical measures capture only partial (several points) information in the yield curve while the three $\hat{\beta}$'s capture the dynamics of entire yield curve.

the level and slope factors $(\hat{\beta}_{1t},\ \hat{\beta}_{2t})$ (denoted CI-NS-Level+Slope) to see whether the level factor or the combination of level and slope factors have dominant contribution in forecasting output growth and inflation.

2.2 Factorizing yield curve information indirectly: CF forecasting schemes

Instead of pooling all the yield information directly into one model, one may first run regressions of y_{t+h} on each element $x_{it} \equiv x_t(\tau_i)$ of x_t to generate individual forecasts $\hat{y}^{(i)}_{T+h} = (1\ x_{iT})\hat{a}_{i,T}$ $(i = 1, 2, \ldots, N)$ and then combine them through some weighting methods. We consider various combination weights. Two simple CF methods are to take equal weights (denoted CF-Mean) to compute $\hat{y}^{\text{CF-Mean}}_{T+h} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}^{(i)}_{T+h}$, and to take the median of the set of N individual forecasts (denoted CF-Median). Besides these simple weighting methods, the combination weights w_i may be estimated from the data to explore more cross-sectional information from the set of individual forecasts. These CF methods are CF-RA, CF-PC, and CF-NS, which we discuss below. Among them, CF-NS is the new method introduced in this paper.

CF-RA: Granger and Ramanathan (1984) suggest to estimate the combination weights w_i by *R*egression *A*pproach:

$$y_{t+h} = w_0 + \sum_{i=1}^{N} w_i \hat{y}^{(i)}_{t+h} + e_{t+h}. \tag{6}$$

As the individual forecasts $\hat{y}^{(i)}_{T+h}$ are constructed to be unbiased, we form the CF-RA predictor with zero intercept: $\hat{y}^{\text{CF-RA}}_{T+h} = \sum_{i=1}^{N}\hat{w}_{i,T}\hat{y}^{(i)}_{T+h}$.

When N is large, the noise in estimating combination weights can dominate any potential improvement (see Min and Zellner 1993), hence we also consider shrinkage weights based on CF-RA to reduce the noise. Let CF-RA(κ) denote the shrinkage forecasts considered in Stock and Watson (2004, p. 412) (see also Aiolfi and Timmermann 2006) with the shrinkage parameter κ controlling for the amount of shrinkage on CF-RA towards the equal weighting (CF-Mean). The shrinkage weight used is $w_{i,T} = \theta\hat{w}_{i,T} + (1-\theta)/N$ with $\theta = \max\{0, 1-\kappa N/(T-h-T_0-N)\}$, where N is the number of individual forecasts, and T_0 is the time when the first pseudo out-of-sample forecast is generated.[2] For simplicity we consider a spectrum of different values of κ, that are chosen such that CF-RA(κ) for the largest chosen value of κ is closest to CF-Mean. For space we report for $\kappa = 0, 1$ in Tables 2 and 3.

[2] In our empirical study, we compute the out-of-sample forecasts by rolling window scheme with in-sample estimation window of size R, and choose T_0, the time when the first pseudo out-of-sample forecast is generated, at the middle point of each rolling window so that it moves along with rolling windows as T is moving towards the end of the out-of-sample period.

CF-PC: Further, for reasons similar to those for CI-PC, we consider factorizing the vector of individual forecasts $\hat{Y}_{t+h} \equiv (\hat{y}^{(1)}_{t+h},\ \hat{y}^{(2)}_{t+h},\ \ldots,\ \hat{y}^{(N)}_{t+h})'$ by extracting their principal components and form predictor CF-PC (combination of forecasts with principal component approach). See Chan, Stock and Watson (1999), Stock and Watson (2004), and Huang and Lee (2006). Specifically we decompose the covariance matrix of these individual forecasts $\hat{Y}_{t+h}$ as $Q\Lambda Q'$, where the diagonal elements of Λ are the eigenvalues and the columns of Q are the associated eigenvectors. Denote the largest k eigenvalues by λ_1, λ_2, ..., and λ_k, and denote the associated eigenvectors by q_1, q_2, ..., and q_k. The first k principal components of $\hat{Y}_{t+h}$ are then defined by $\hat{F}^{(i)}_{t+h} = q_i'\hat{Y}_{t+h},\ i = 1, 2, \ldots, k$. From using each $\hat{F}^{(i)}_{t+h}, i = 1, 2, \ldots, k$ in a regression of y_{t+h} on it, we form a PC combined forecast (denoted as CF-PC (1st), CF-PC (2nd), and so on, respectively) where the weights are proportional to the eigenvector q_i at each out-of-sample point in time.

We can further combine k_0 of these PC combined forecasts to produce additional forecasts. First we estimate

$$y_{t+h} = \sum_{i=1}^{k_0} b_i \hat{F}^{(i)}_{t+h} + u_{t+h}, \tag{7}$$

for $t = T_0, \ldots, T$.[3] From this equation we construct a forecast (denoted as CF-PC($k = k_0$)) as $\hat{y}^{\text{CF-PC}}_{T+h} = \sum_{i=1}^{k_0} \hat{b}_{i,T} \hat{F}^{(i)}_{T+h}$. The number of principal components k_0, can be selected either by information criteria such as AIC and BIC, or fixed at small values like 1, 2, or 3.

CF-NS: Finally, motivated by the principle of parsimony, we apply the NS exponential factorizing framework to CF, and call it CF-NS (combination of forecasts with NS exponential factorizing framework). Note that while the CF-PC method is suited for data of many kinds, the CF-NS method we propose is tailored to forecasting using yield curve. It uses fixed factor loadings

$$1,\ \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i}\right),\ \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right),$$

that are the NS exponential factor loadings for yield curve distillation, and hence avoids the estimation of factor loadings (as CF-PC does) that may potentially deteriorate out-of-sample forecasting performance in case of large noise in the data (the fixed factor loadings may be interpreted as another type of shrinkage, similar to using fixed equal weights in CF-Mean). The procedure of CF-NS is as follows. First, we note that after normalization, these three sets of fixed weights (the

[3] We do not include an intercept term in the regression model because each estimated principal component may be regarded as already unbiased since they are linear combinations of individual forecasts which are bias-adjusted by construction. In addition, we find that excluding intercept term generally helps us achieve better forecasting performance than the one including intercept term.

three NS exponential loadings) can be used as forecast combination weights. Therefore, we form three NS combined forecasts as follows:

$$\begin{aligned} z_{1,t+h} &= \frac{1}{s_1}\sum_{i=1}^{N}\hat{y}_{t+h}^{(i)} \\ z_{2,t+h} &= \frac{1}{s_2}\sum_{i=1}^{N}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}\right)\hat{y}_{t+h}^{(i)} \\ z_{3,t+h} &= \frac{1}{s_3}\sum_{i=1}^{N}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}-e^{-\lambda\tau_i}\right)\hat{y}_{t+h}^{(i)} \end{aligned}$$

where $s_1 \equiv \sum_{i=1}^{N} 1 = N$, $s_2 \equiv \sum_{i=1}^{N}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}\right)$, and $s_3 \equiv \sum_{i=1}^{N}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}-e^{-\lambda\tau_i}\right)$.[4] Note that by construction, $z_{1,T+h} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_{t+h}^{(i)}$ is the CF-Mean. In Figure 1 we plot the three normalized NS exponential loadings, which shows that the first combined forecast $z_{1,t+h}$ has equal weights on all individual forecasts $\hat{y}_{t+h}^{(i)}$ and thus captures the yield curve level factor (may be denoted as CF-NS-Level, which is the same as CF-Mean); the second combined forecast $z_{2,T+h}$ has decreasing weights assigned to individual forecasts using yields with increasing maturities and thus captures the yield curve slope factor and is denoted as CF-NS-Slope; and the third combined forecast $z_{3,T+h}$ has first increasing then decreasing weights for individual forecasts from yields with increasing maturities and thus captures the yield curve curvature factor and is therefore denoted as CF-NS-Curvature. The CF-NS-Slope and CF-NS-Curvature methods are examined in order to see the separate contributions of the NS combined forecasts that capture slope or curvature factor only.

To see how the three NS combined forecasts contribute to forecasting the variable of interest y_{t+h} *jointly*, one can pool these three NS combined forecasts by regression:

$$y_{t+h} = c_1 z_{1,t+h} + c_2 z_{2,t+h} + c_3 z_{3,t+h} + v_{t+h} \tag{8}$$

to get $\hat{y}_{T+h}^{\text{CF-NS}} = \sum_{i=1}^{3}\hat{c}_{i,T} z_{i,T+h}$. In the empirical section, the two CF-NS forecasts considered are: CF-NS-All with no restrictions on c's and CF-NS-Level+Slope with the restriction $c_3 = 0$ in equation (8).

Additionally, with the $N = 17$ individual forecasts we have in our empirical study, instead of using NS fixed loadings to combine them into CF-NS-All, we could use only the individual forecasts based on 3-month, 2-year, and 10-year yields (components of yield curve empirical level, slope, and curvature measures (Diebold and Li 2006) and use regression to combine these three forecasts only (denoted CF-Empirical-Measures) to avoid the large-N problem at the same time accounting for

[4] In our empirical study, $(\tau_1\ \tau_2\ \ldots\ \tau_N) = (3\ 6\ 9\ 12\ 15\ 18\ 21\ 24\ 30\ 36\ 48\ 60\ 72\ 84\ 96\ 108\ 120)$, the seventeen maturities in months.

representative information in the yield curve. However, this method is ignoring information in other yields so its performance is found to be worse than CF-NS-All in our empirical analysis. To incorporate the entire yield curve information through CF, alternatively one can combine forecasts from the three NS factors ($\hat{\beta}$'s), i.e., using $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ separately to generate three individual forecasts and then combining them with a regression (denoted CF-NS-Factors). Performance of CF-NS-Factors is very close to that of CF-Empirical-Measures, but not necessarily better because the three NS factors are estimated even though they capture more information in the yield curve. That CF-NS-All is generally better than these two alternatives is probably due to the argument we elaborate in the next section (Section 3.4.7).

3 Forecasting Output Growth and Inflation: Empirics

This section presents the empirical analysis where we describe the data, implement forecasting methods introduced in the previous section on forecasting output growth and inflation, and analyze out-of-sample forecasting performances. This allows us to draw the differences between output growth and inflation forecasting using the same yield curve information and to compare the strengths of different methods.

3.1 Data

Two forecast targets, output growth and inflation, are constructed respectively as the monthly growth rate of Personal Income (PI, seasonally adjusted annual rate) and monthly change in CPI (Consumer Price Index for all urban consumers: all items, seasonally adjusted) from 02/1970 to 09/2005. PI and CPI data are obtained from the website, http://research.stlouisfed.org/fred2/ of Federal Reserve Bank at St. Louis. In Section 2, we use y_{t+h} to denote the variable to be forecast (output growth or inflation) using the yield information $\mathbf{x}_t$ up to time t. y_{t+h} may be obtained after proper transformation of the original variables (PI or CPI). For instance, when forecasting the monthly growth rate of PI, we set $y_{t+h} = 1200[(1/h)\ln(\text{PI}_{t+h}/\text{PI}_t)]$ as the forecast target (as used in Ang, Piazzesi and Wei (2006)). For the consumer price index (CPI), we set $y_{t+h} = 1200[(1/h)\ln(\text{CPI}_{t+h}/\text{CPI}_t)]$ as the forecast target (as used in Stock and Watson (2005b)).

Our yield curve data consist of U.S. government bond prices, coupon rates, and coupon structures, as well as issue and redemption dates. We calculate zero-coupon bond yields using the unsmoothed Fama-Bliss (1987) approach. We measure the bond yields on the second day of each month. We also apply several data filters designed to enhance data quality and focus attention on maturities with good liquidity. First, we exclude floating rate bonds, callable bonds and bonds extended beyond the original redemption date. Second, we exclude outlying bond prices less than 50 or greater than 130 because their price discounts/premium are too high and imply thin trading, and we exclude yields that differ greatly from yields at nearby maturities. Finally, we use only bonds with maturity greater than one month and less than fifteen years, because other bonds are not actively traded. Indeed, to simplify our subsequent estimation, using linear interpolation we pool the bond yields into fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, where a month is defined as 30.4375 days.

In Table 1 we provide descriptive statistics of the two forecast targets and yield curve level, slope, and curvature (empirical measures), over three different sample periods: full sample from 02/1970 to 09/2005, the first out-of-sample evaluation period from 01/1985 to 09/2005, and the second out-of-sample evaluation period from 01/1995 to 09/2005. The level is defined as the 10-year yield $x_t(120)$, the slope as the difference between the 10-year and 3-month yields $x_t(120) - x_t(3)$, and the curvature as the twice the 2-year yield minus the sum of the 3-month and 10-year yields $2x_t(24) - (x_t(3) + x_t(120))$. It is clear that both PI growth and CPI inflation become more moderate and less volatile after 1985. This stylized fact is known as "Great Moderation". See Kim and Nelson (1999) and D'Agostino *et al.* (2005). In particular, there is a substantial drop in the persistency of CPI inflation over the period of 1985 to 2005. The volatility and persistency of yield curve slope and curvature do not change much. Yield curve level, however, decreases and stabilizes a lot after 1985 and these trends continue after 1995.

3.2 The choice of yield levels instead of spreads

In predicting macroeconomic variables using the term structure, yield spreads between yields with various maturities and the short rate are commonly used in the literature. One possible reason for such a practice is that yield levels are treated as I(1) processes so yield spreads will likely be I(0). Similarly macroeconomic variables are typically assumed as I(1) and transformed properly into I(0) so that when using yield spreads to forecast macro targets, issues such as spurious regression are avoided. In this paper, however, we use yield levels (not spreads) to predict PI growth and CPI inflation (not change in inflation). First, whether yields and inflation are I(1) or I(0) is still arguable. Stock and Watson (1999, 2005a,b) use yield spreads and treat inflation as I(1) so they forecast change in inflation. Inoue and Kilian (2006), however, treat inflation as I(0). Since our target is forecasting inflation, not change in inflation, we will treat CPI inflation as well as yields as I(0) in our empirical analysis. Second, we emphasize real-time, out-of-sample forecasting performance more than in-sample concerns. As long as out-of-sample forecast performance is unaltered or even improved, we think the choice of treating them as I(1) or I(0) variables does not matter much.[5] Third, using yield levels will allow us to provide clearer interpretations for questions such as, which part of the yield curve contributes the most towards predicting PI growth or CPI inflation, and how the different parts of yield curve interact in the prediction, etc.

3.3 Out-of-sample forecasting

We now turn to the specifics related to our forecasting methods introduced in Section 2 and examine their out-of-sample performance. All forecasting models are estimated by rolling scheme with window size R. We consider two different out-of-sample evaluation periods: one from 01/1985 to 09/2005 (hence out-of-sample size $P = 249$ and $R = 179$), and another more recent one from 01/1995 to 09/2005 ($P = 129$ and $R = 299$). In all PC related forecasting methods, we choose the hypothesized upper limit for the true number of factors, k_{max}, at seventeen (the total number of yield levels used). In all NS related methods (for both CI and CF) we set λ, the parameter that governs the exponential decay rate, at 0.0609 for reasons as in Diebold and Li (2006).[6]

We compare h-month-ahead out-of-sample forecasting results of those methods introduced in Section 2 with two simple benchmark models, for $h = 1, 3, 6$, and 12 months. For both PI growth and CPI inflation, the first benchmark is the AR model where $\hat{y}_{t+h} = \hat{\zeta}_{0,t} + \hat{\zeta}_{1,t} y_t$.[7] For PI growth, the second benchmark is the model from Ang, Piazzesi and Wei (2006) estimated by OLS (denoted APW-OLS):

$$y_{t+h} = \phi_0 + \phi_1 x_t(3) + \phi_2 (x_t(60) - x_t(3)) + \phi_3 y_t + \varepsilon_{t+h}. \quad (9)$$

For CPI inflation, the second benchmark is the Stock and Watson's (2005b) IMA(1,1) univariate model with its moving average coefficient estimated using a ten-year rolling window of past observations (denoted IMA(1,1)-10-year).

3.4 Results

Tables 2 and 3 present the root mean squared forecast errors (RMSFE) of all methods for PI growth (Table 2) and CPI inflation (Table 3) forecasts using all seventeen yield levels. These results are

[5] While not reported for space, we tried forecasting change in inflation and found forecasting inflation directly using all yield levels improves out-of-sample performances of most forecasting methods by a large margin.

[6] For different values of λ, the performances of CI-NS and CF-NS change only marginally.

[7] We use the direct multistep AR method rather than the iterated multistep AR method, as all the other models discussed in Section 2 and the second benchmark model are based on the direct forecasts that are made using a horizon-specific estimated model where the dependent variable is the multiperiod ahead value being forecast. The direct forecasts may be more robust to model misspecification. See Marcellino, Stock, and Watson (2006) for more discussion.

summarized as follows.

3.4.1 CF is better than CI.

In most cases, the benchmark models are clearly outperformed either by CF-PC, or CF-NS, or sometimes CF-RA(κ). The improvements over benchmarks by these methods are sometimes quite substantial. In contrast, when forecasting PI growth, only in few cases the CI-PC or CI-NS can beat the benchmark models just by a small margin, and when forecasting CPI inflation, in almost no cases the benchmarks are beaten by CI schemes with most of them (including CI-PC and CI-NS) performing much worse than the benchmarks. Both CI-PC and CI-NS are worse than CF models. This finding, the superior performance of CF in comparison with CI in real-time forecasting, is consistent with what analytically and empirically demonstrated in Huang and Lee (2006). While the performance of CF-RA(κ) largely depends upon the choice of κ (the parameter controlling the amount of shrinkage towards equal weights), CF-PC and CF-NS behave well as long as more than one factor are included.

In using information in the entire yield curve $\mathbf{x}_t = (x_{1t} \dots x_{Nt})'$ to forecast y_{t+h}, CF combines individual forecasts $\hat{y}_{t+h}^{(i)}$ each obtained from using one yield x_{it} at a time or combines particular combinations of some yields in $\mathbf{x}_t$, while CI uses the entire yield information $\mathbf{x}_t$ in one big model. When N is large, both CI and CF without using factorization would suffer from parameter estimation error, because α and w_i's need to be estimated for $\hat{y}_{T+h}^{\text{CI-Unrestricted}} = (1\ \mathbf{x}_T')\hat{\alpha}_T$ and $\hat{y}_{T+h}^{\text{CF-RA}} = \sum_{i=1}^{N} \hat{w}_{i,T}\hat{y}_{T+h}^{(i)}$. We have considered factor models (PC or NS) to reduce the dimension of $\hat{Y} \equiv (\hat{y}^{(1)} \dots \hat{y}^{(N)})'$ for CF and the dimension of $\mathbf{x} \equiv (x_1 \dots x_N)'$ for CI.

From our empirical results we find that CF (CF-PC and CF-NS) using the factors of forecasts $(\hat{y}^{(1)} \dots \hat{y}^{(N)})$ is better than CI (CI-PC and CI-NS) using the factors of the yields $(x_1 \dots x_N)$. The main reason is that CF incorporates the "relationship" between the forecast target y and predictors x_i's as the factors are extracted from $(\hat{y}_1 \dots \hat{y}_N)$ after each individual forecast is obtained from incorporating the relationship between y and each x_i, while CI extracts the factors of $\mathbf{x}_t$ without taking their relationship with the forecast target y into account. Therefore, in spirit, CF-PC and CF-NS are similar to the "partial least squares" method. See, e.g., Garthwaite (1994).

3.4.2 CF-NS is more stable than CF-PC.

CF-PC(2nd) and CF-PC(3rd) are poor performers. This indicates that individual PCs (the second and third one) have bad weights so that they do not work for prediction while the first PC (about equal weights) works. These can be observed for both output growth (Table 2) and inflation (Table

3) forecasting results. Looking at the factor loadings in Figure 2 where we plot the average of three factor loadings associated with the first, second and third principal components in CF-PC, the second PC and third PC assign negative weights for some individual forecasts and weights bigger than unity for others. In contrast, the weights used in CF-NS are all between zero and one (see Figure 1) as they are normalized.[8]

3.4.3 Slope and curvature factors are important for longer horizon forecasting.

The slope and curvature factors play important role in improving forecasting performance of the yield curve for both output growth and inflation, especially for long horizon forecasting ($h = 3, 6, 12$) and more so as h gets larger.

3.4.4 The number of factors needed depends on the forecast target and the forecast horizon h.

For one month-ahead ($h = 1$) forecasting of the PI growth, generally the best method is CF-PC with only one factor. For the CPI inflation forecasting, for better performance generally we need two or three factors in CF-PC models. This is also true with CF-NS models. Figure 2 helps us understand what economic contents these factors in CF-PC may bear. It shows that the first PC assigns about equal weights to all seventeen individual forecasts that use yields at various maturities (in months) so that it may be interpreted as the factor that captures the level of the yield curve; the second PC assigns roughly increasing weights so that it may be interpreted as factor capturing the slope; and the third PC assigns roughly first decreasing then increasing weights so that it may be interpreted as factor capturing the curvature. Hence it seems that output growth forecasting requires only one factor that captures the level of the yield curve for $h = 1$, while it requires two or three factors that captures the level, slope and curvature of the yield curve for $h > 1$.[9] When $h = 6$ or 12, CF-NS-Level+Slope and CF-NS-All perform much better than CF-NS-Level (CF-Mean) for PI growth forecasting (Table 2). This is also somewhat true for CPI inflation forecasting in Table 3. Both output growth and inflation forecasting generally require two or three factors that captures

[8] We also experimented with the normalized factor loadings for the CF-PC method, which however turned out to have quite erratic performances.

[9] For output growth, Kozicki (1997) finds that the predictive power of the yield spread largely comes from its usefulness over horizons of a year or so and generally dominates the predictive power associated with the level of yields. This is similar to what we find. However, for inflation, she finds that although the yield spread helps predict inflation at moderate horizons of a few years, the level of yields is a more useful predictor of inflation. We note that Kozicki's results are based on the in-sample analysis while ours are on out-of-sample forecasting. She uses quarterly sample from 1970 to 1996.

the level, slope and curvature of the yield curve for $h > 1$.[10]

3.4.5 CF-NS is more successful than IMA(1,1) especially in longer horizon inflation forecasts.

In forecasting the CPI inflation, although IMA(1,1) is the best model for shorter horizons, we can beat it by either CF-PC or CF-NS (mostly CF-NS) in longer horizons. Note that in the first out-of-sample period starting from 1985, CF-NS-All appears to be generating the best forecast and in the second out-of-sample period starting from 1995, CF-NS-Level+Slope performs generally the best among all CF-NS methods. To further evaluate the relative out-of-sample performance of these two competing forecasts, we consider the following unconstrained forecast-encompassing regression (as in Harvey and Newbold (2005)):

$$y_{t+h} = d_0 + d_1 \hat{y}^{\text{IMA}}_{t+h} + d_2 \hat{y}^{\text{CF-NS}}_{t+h} + u_{t+h}, \tag{10}$$

where $\hat{y}^{\text{IMA}}_{t+h}$ denotes the forecast by IMA(1,1) model, and $\hat{y}^{\text{CF-NS}}_{t+h}$ in the first and second out-of-sample period are forecasts by CF-NS-All and CF-NS-Level+Slope, respectively. Table 4 presents the estimated regression coefficients (with Newey and West (1987) standard errors in parenthesis) from regressing CPI inflation y_{t+h} on the two competing forecasts plus a constant in two out-of-sample periods and for $h = 1, 3, 6,$ and 12. Generally the results are consistent with the RMSFE results in Table 3: CF-NS-All forecast (in the first out-of-sample period) or CF-NS-Level+Slope (in the second out-of-sample period) gain bigger weights and are more significant as forecast horizon h grows. From the "forecast-encompassing" (Chong and Hendry 1986) point of view, in case when we fail to reject $H_0 : d_1 = 0$, we may conclude that the CF-NS forecast encompasses IMA(1,1). Table 4 shows that it seems in most cases neither one encompasses the other but as forecast horizon h grows, CF-NS is more capable of forecast-encompassing IMA(1,1).

3.4.6 In forecasting inflation, the predictive power of the yield curve for the period 1995-2005 is weaker than for the period 1985-2005.

This is consistent with recent flat/inverted yield curve but no recession. The results in Table 3 show that it appears to be harder to beat IMA(1,1) in more recent period since for out-of-sample period starting from 1995 we can outperform IMA(1,1) only at 1-year horizon while for the one starting from 1985, we can beat it at 3, 6, and 12-month forecast horizons. This is probably due to the

[10] Wright (2006) finds some similar results in forecasting turning point of the business cycles. He finds that models that use both the level and the slope give better in-sample fit and out-of-sample predictive performance (in forecasting recessions using the yield curve) than models with the slope alone.

"Great Moderation" during which one can hardly improve upon naive univariate models and macro and financial variables have less predictive power towards predicting inflation (see D'Agostino, Giannone, and Surico 2005). Our finding may indicate diminishing forecasting power of the yield curve within the Great Moderation period. This may be explained by the views in Estrella (2005, p. 742): "In fact, ... the estimates shed some light on the connection between monetary policy and the varying predictive relationship between the yield spread, output and inflation. Particularly notable are the estimates in the post-1987 period, which seems to be consistent with strict inflation targeting and in which the predictive power of the yield spread, though not entirely absent, is certainly diminished."

3.4.7 CF-NS-All is better than CF-Mean.

In general, the new method (namely, CF-NS) that we introduce in this paper is the best way to extract level, slope and curvature information from the yield curve for forecasting output growth and inflation. The yield curve factors that can not be captured by one single individual forecast may be well captured by CF methods since they pool all the seventeen individual forecasts together hence incorporate the entire yield curve information. More importantly, we find here that CF-NS is better than CF-Mean, due to the fact that slope and curvature information of yield curve matters. To be more specific, all the seventeen yields matter for predicting the output growth and inflation, but they do not contribute evenly. This uneven contribution, which is bypassed in CF-Mean (i.e., CF-NS-Level), could be uncovered through other CF-NS methods since in forming the combination weights they include factors that assign unequal weights to individual forecasts thus capturing the slope and/or curvature (Figure 1) information. The ability of these CF-NS methods to capture parsimoniously the slope and curvature information in the yield curve attributes to its superior out-of-sample forecasting performance for output growth and inflation.

This finding that CF-NS is better than CF-Mean is in contrast to the general findings in the literature that CF-Mean is often found to be the best (see, for example, Timmermann 2005). Stock and Watson (2004) call this a "forecast combination puzzle" that the simple combinations as in CF-Mean are repeatedly found to outperform sophisticated/adaptively weighted combinations in empirical applications. In particular, Smith and Wallis (2005) explores a possible explanation of the forecast combination puzzle. Their explanation lies in the effect of finite sample estimation error of the combining weights. However, in our paper, we show cases where using more sophisticated combining weights as in CF-NS can be clearly better than using the simple weights as in CF-Mean. We now investigate how this may happen.

Note that CF-NS-All ($\hat{y}_{T+h}^{\text{CF-NS-All}} = \sum_{i=1}^{3} \hat{c}_{i,T} z_{i,T+h}$) is obtained from the regression of equation (8) with no restrictions on c's, CF-NS-Level+Slope ($\hat{y}_{T+h}^{\text{CF-NS-Level+Slope}} = \sum_{i=1}^{2} \hat{c}_{i,T} z_{i,T+h}$) is obtained with the restriction $c_3 = 0$, and CF-Mean ($\hat{y}_{T+h}^{\text{CF-Mean}} = z_{1,T+h}$) is obtained with restrictions $c_1 = 1, c_2 = c_3 = 0$. We also consider three other methods of calculating $\hat{c}_{i,T}$ for CF-NS-All, as reported in Tables 2 and 3: namely, CF-NS-All-Mean (with $\hat{c}_{i,T} = \frac{1}{3}$ for $i = 1, 2, 3$), CF-NS-All-BMA (with Bayesian model averaging weights as discussed in Lee and Yang (2006)), and CF-NS-All-Yang (with weights of Yang (2004)). From Tables 2 and 3, we observe that CF-NS-All-Mean, CF-NS-All-BMA, and CF-NS-All-Yang are almost the same as CF-Mean, and they are clearly worse than CF-NS-Level+Slope and CF-NS-All. Therefore the result that CF-NS is better than CF-Mean depends on how we obtain c's to form CF-NS. In order to understand it better, we look at the estimated weights ($\hat{c}_{1,T}$ $\hat{c}_{2,T}$) for CF-NS-Level+Slope reported in Figure 3 and ($\hat{c}_{1,T}$ $\hat{c}_{2,T}$ $\hat{c}_{3,T}$) for CF-NS-All in Figure 4. The large variation in $\hat{c}_{i,T}$ and the mirror-image behavior between $\hat{c}_{i,T}$'s indicate the collinearity between z's. The collinearity between z's is due to the collinearity between the individual forecasts ($\hat{y}^{(1)} \ldots \hat{y}^{(N)}$) which is likely caused by the high correlation among x_{it}'s (yields at different maturities) that are used to generate the individual forecasts.

This collinearity, however, is not necessarily damaging. Analogous to the "pairs trading" strategy (see, for example, Gatev, Goetzmann, and Rouwenhorst (2006)) which trades on two highly correlated (or cointegrated) stocks whose prices have moved together historically, CF-NS utilizes a set of highly correlated/collinear forecasts. In this case, the optimal forecast combination should include the worse forecasts with negative weights (as short-sale of one in pairs trading) and the better forecasts with weights larger than 1 (as long the other in pairs trading). This can be achieved by the regression approach (such as CF-NS-Level+Slope or CF-NS-All), but not by the methods restricting the combining weights on the (0 1) interval (such as CF-Mean, CF-NS-All-Mean, CF-NS-All-BMA, and CF-NS-All-Yang).

Bates and Granger (1969) consider the case of combining two unbiased one-step ahead forecasts. Let $f_t^{(1)}$ and $f_t^{(2)}$ be forecasts of y_{t+1} with errors

$$e_{t+1}^{(j)} = y_{t+1} - f_t^{(j)}, \qquad j = 1, 2$$

such that $Ee_{t+1}^{(j)} = 0$, $Ee_{t+1}^{(j)2} = \sigma_j^2$, and $Ee_{t+1}^{(1)} e_{t+1}^{(2)} = \rho\sigma_1\sigma_2$. A combined forecast

$$f_t^{(c)} = kf_t^{(1)} + (1-k)f_t^{(2)}$$

has the forecast error $e_{t+1}^{(c)} = y_{t+1} - f_t^{(c)} = ke_{t+1}^{(1)} + (1-k)e_{t+1}^{(2)}$ and the forecast error variance

$\sigma_c^2 = k^2\sigma_1^2 + (1-k)^2\sigma_2^2 + 2k(1-k)\rho\sigma_1\sigma_2$, which is minimized for the value of k given by

$$k_{opt} = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

The optimal weight k_{opt} yields the minimum error variance $\sigma_{c,opt}^2 = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$, for which $\sigma_{c,opt}^2 < \min(\sigma_1^2, \sigma_2^2)$.

The situation $k_{opt} < 0$ is interesting. In light of the above condition, it appears that an inferior forecast may still be worth including with negative weight. This happens when $\sigma_2^2 - \rho\sigma_1\sigma_2 < 0$ or $\sigma_2/\sigma_1 < \rho$, i.e., when ρ is a very large positive value, say close to 1, and $f_t^{(1)}$ is the inferior forecast with larger forecast error variance σ_1.

As shown in Granger and Newbold (1986, p. 268), the optimal combining weight k_{opt} can be estimated from

$$\hat{k}_t = \frac{\sum_{s=1}^{t}\left(e_s^{(2)2} - e_s^{(1)}e_s^{(2)}\right)}{\sum_{s=1}^{t}\left(e_s^{(1)2} + e_s^{(2)2} - 2e_s^{(1)}e_s^{(2)}\right)}$$

which can be obtained from the regression $e_{t+1}^{(2)} = k\left(e_{t+1}^{(2)} - e_{t+1}^{(1)}\right) + e_{t+1}^{(c)}$.

A common popular recommendation is to ignore ρ. For example, Clemen (1989, p. 562) suggests "to ignore the effect of correlations in calculating combining weights". While the optimal weight $\hat{k}_t$ can be negative or overweighted (larger than one) depending on the value of ρ, the use of a simpler form obtained with the restriction $\rho = 0$ has been a popular recommendation:

$$\hat{k}_t' = \frac{\sum_{s=1}^{t} e_s^{(2)2}}{\sum_{s=1}^{t}\left(e_s^{(1)2} + e_s^{(2)2}\right)}.$$

Note that ignoring ρ, $\hat{k}_t'$ is always constrained on the (0 1) interval (analogous to the short-sale constraint). Examples of weights with this constraint include equal weights (as in CF-NS-All-Mean and CF-Mean), BMA weights (as in CF-NS-All-BMA), and Yang's (2004) weights (as in CF-NS-All-Yang).

When ρ is large and positive, the optimal weight on the inferior forecast can be negative. The forecast combination problem is analogous to that of minimizing the variance of a portfolio, with the forecast errors playing the role of asset returns (Timmermann 2005). Gatev, Goetzmann, and Rouwenhorst (2006) show that the "pairs trading" in financial trading strategy profits from the high correlation in the returns. Analogously, the profitability of using the optimal weight is linked to the high correlation ρ in the forecasts. Without loss of generality lets assume $f_t^{(1)}$ is the inferior forecast with larger forecast error variance. In combining forecasts, when $\rho \gg 0$, we short the loser (the worse forecast) with $k < 0$ and buy the winner (the better one) with $(1-k) > 1$.

Figure 5 illustrates how the correlation ρ (similarity of forecasts) matters in the calculation of the forecast combination weights. When $\rho < 0$ (two individual forecasts are distinct and fall on the different sides of y on the U curve of the MSFE loss function) as shown in Figure 5(a), the optimal weight $\hat{k}_t$ on the inferior forecast $f_t^{(1)}$ falls on (0 1) interval and therefore using $\hat{k}_t'$ may not lose much and can work well. When $\rho > 0$ (two individual forecasts are similar and fall on the same side of y) as shown in Figure 5(b), the optimal weight $\hat{k}_t$ on the inferior forecast $f_t^{(1)}$ should be negative for combined forecast to achieve smaller loss than both individual ones. In this case, the use of $\hat{k}_t'$ ignoring the correlation would be too restrictive. Therefore, when the forecasts $(\hat{y}^{(1)}, \ldots, \hat{y}^{(N)})$ are collinear thus resulting highly correlated and similar z's, CF-NS-All can be better than CF-Mean since it accounts for the collinearity of z's by assigning the optimal (non-convex) weights (that are negative or greater than one).

4 Conclusions

We propose a new forecasting method for forecasting the macro-variables using the entire yield curve (which we term as CF-NS), that applies the Nelson-Siegel yield curve factorizing framework to combination of forecasts. The CF-NS method is first to combine forecasts from individual yields with three sets of fixed weights that are the three normalized Nelson-Siegel exponential factor loadings corresponding respectively to level, slope and curvature factors of the yield curve, then to estimate a regression of the variable to be forecast on these three combined forecasts, and finally to form the forecast based upon this regression.

We have found that CF is better than CI and the newly proposed CF-NS method has better out-of-sample performance than several popular benchmark models in forecasting monthly PI growth at all horizons and in forecasting CPI inflation in longer horizons. CF-NS and CF-PC are like the partial least squares in spirit, in that we extract factors to reduce collinearity (resulted from individual forecasts using one yield at a time) and at the same time in a way that accounts for the relationship between the yields $\mathbf{x}_t$ and the forecast target y_{t+h}. As the three Nelson-Siegel combined forecasts (z_1, z_2, z_3) are highly positively correlated and similar to each other, the optimal combination of z's may be with non-convex weights as CF-NS does when pooling z's. This resembles the pairs trading strategy (going short one and long the other) and sheds light on the reasons why CF-NS works better than CF-Mean.

References

Aiolfi, M. and Timmermann, A. (2006), "Persistence in Forecasting Performance and Conditional Combination Strategies," *Journal of Econometrics* 135, 31-53.

Ang, A. and Piazzesi, M. (2003), "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics* 50, 745-787.

Ang, A., Piazzesi, M, and Wei, M. (2006), "What Does the Yield Curve Tell Us about GDP Growth?" *Journal of Econometrics* 131, 359-403.

Bai, J. and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70, 191-221.

Bernanke, B.S. (1990), "On the Predictive Power of Interest Rates and Interest Rate Spreads," Federal Reserve Bank of Boston, *New England Economic Review*, 51-68.

Chan, Y.L., Stock, J.H., and Watson, M.W. (1999), "A Dynamic Factor Model Framework for Forecast Combination," *Spanish Economic Review* 1, 91-121.

Chong, Y.Y. and Hendry, D.F. (1986), "Econometric Evaluation of Linear Macro-Economic Models," *Review of Economics Studies* LIII, 671-690.

Clemen, R.T. (1989), "Combining Forecasts: A Review and Annotated Bibliography", *International Journal of Forecasting,* 5, 559-583.

D'Agostino, A., Giannone, D., and Surico, P. (2005), "(Un)Predictability and Macroeconomic Stability," ECARES, Universite Libre de Bruxelles.

Diebold, F.X. and Li, C. (2006), "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics* 130, 337-364.

Diebold, F.X., Piazzesi, M., and Rudebusch, G.D. (2005), "Modeling Bond Yields in Finance and Macroeconomics," *American Economic Review* 95, 415-420.

Diebold, F.X., Rudebusch, G.D., and Aruoba, B. (2006), "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *Journal of Econometrics* 131, 309-338.

Estrella, A. (2005), "Why Does the Yield Curve Predict Output and Inflation?" *The Economic Journal* 115, 722-744.

Estrella, A. and Hardouvelis, G.A. (1991), "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance* 46, 555-76.

Fama, E. and Bliss, R. (1987), "The Information in Long-maturity Forward Rates," *American Economic Review* 77, 680-692.

Friedman, B.M., and Kuttner, K.N. (1991), "Why Does the Paper-Bill Spread Predict Real Economic Activity?" *NBER Working Paper*, No. 3879.

Garthwaite, P.H. (1994), "An Interpretation of Partial Least Squares,"*Journal of the American Statistical Association* 89, 122-127.

Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G. (2006), "Pairs Trading: Performance of a Relative-Value Arbitrage Rule," *Review of Financial Studies,* 19(3), 797-827.

Granger, C.W.J. and P. Newbold (1986), *Forecasting Economic Time Series*, 2ed., Academic Press.

Granger, C.W.J. and Ramanathan, R. (1984), "Improved Methods of Combining Forecasts," *Journal of Forecasting* 3, 197-204.

Hamilton, J.D. and Kim, D.H. (2002), "A Reexamination of the Predictability of Economic Activity Using the Yield Spread," *Journal of Money, Credit, and Banking* 34, 340-360.

Harvey, D.I. and Newbold, P. (2005), "Forecast Encompassing and Parameter Estimation," *Oxford Bulletin of Economics and Statistics* 67(Supplement), 815-835.

Huang, H. and Lee, T.-H. (2006), "To Combine Forecasts or to Combine Information?" University of California, Riverside.

Inoue, A. and Kilian, L. (2006), "How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation," University of British Columbia and University of Michigan.

Kim, C.-J. and Nelson, C.R. (1999), "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics* 81, 608-616.

Kozicki, S. (1997), "Predicting Real Growth and Inflation with the Yield Spread," Federal Reserve Bank of Kansas City, *Economic Review* 82, 39-57.

Lee, T.-H. and Yang, Y. (2006), "Bagging Binary and Quantile Predictors for Time Series, "*Journal of Econometrics* 135, 465-497.

Marcellino, M., Stock, J.H., and Watson, M.W. (2006), "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics* 135, 499-526.

Min, C. and Zellner, A. (1993), "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," *Journal of Econometrics* 56, 89-118.

Nelson, C.R. and Siegel, A.F. (1987), "Parsimonious Modeling of Yield Curves," *Journal of Business* 60, 473-489.

Newey, W.K. and West, K.D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Piazzesi, M. (2005), "Bond Yields and the Federal Reserve," *Journal of Political Economy*, forthcoming.

Rudebusch, G.D. and Wu, T. (2004), "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," Working paper, Federal Reserve Bank of San Francisco.

Smith, J. and Wallis, K.F. (2005), "Combining Point Forecasts: The Simple Average Rules, OK?" University of Warwick.

Stock, J.H. and Watson, M.W. (1989), "New Indexes of Coincident and Leading Indicators," *NBER Macroeconomic Annual*, Vol. 4, Olivier Blanchard and Stanley Fischer (ed.). Cambridge: MIT Press.

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

Stock, J.H. and Watson, M.W. (2002), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association* 97, 1167-1179.

Stock, J.H. and Watson, M.W. (2004), "Combination Forecasts of Output Growth in a Seven-country Data Set," *Journal of Forecasting* 23, 405-430.

Stock, J.H. and Watson, M.W. (2005a), "An Empirical Comparison of Methods for Forecasting Using Many Predictors," Harvard University and Princeton University.

Stock, J.H. and Watson, M.W. (2005b), "Has Inflation Become Harder to Forecast?" Harvard University and Princeton University.

Timmermann, A. (2005), "Forecast Combinations," forthcoming in *Handbook of Economic Forecasting*, edited by Graham Elliott, Sir Clive Granger and Allan Timmermann, North Holland.

Wright, J.H. (2006), "The Yield Curve and Predicting Recessions," Federal Reserve Board, Washington D.C., Finance and Economics Discussion Series No 2006-07.

Yang, Y. (2004), "Combining Forecasting Procedures: Some Theoretical Results", *Econometric Theory*, 20, 176-222.

Table 1. Descriptive Statistics

		Mean	Std. dev.	Min.	Max.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
Forecast Targets								
	02/1970-09/2005	7.145	7.466	-48.692	43.468	-0.088	0.252	0.084
PI Growth	01/1985-09/2005	5.329	7.923	-48.692	43.468	-0.281	0.285	0.026
	01/1995-09/2005	5.055	6.813	-31.401	43.468	-0.119	0.034	-0.023
	02/1970-09/2005	4.643	3.760	-6.581	21.525	0.642	0.410	0.159
CPI Inflation	01/1985-09/2005	3.046	2.460	-6.581	14.597	0.310	-0.012	0.006
	01/1995-09/2005	2.600	2.508	-4.049	14.597	0.140	-0.031	-0.021
Yield Curve								
	02/1970-09/2005	7.604	2.303	3.349	14.925	0.978	0.768	0.520
Level	01/1985-09/2005	6.686	1.703	3.349	11.663	0.949	0.535	0.387
	01/1995-09/2005	5.414	0.903	3.349	7.742	0.860	0.342	0.023
	02/1970-09/2005	1.466	1.450	-3.505	4.060	0.930	0.417	-0.116
Slope	01/1985-09/2005	1.854	1.247	-0.752	4.060	0.949	0.412	-0.070
	01/1995-09/2005	1.559	1.242	-0.752	3.998	0.938	0.445	0.057
	02/1970-09/2005	-0.054	0.842	-2.375	3.169	0.845	0.380	0.021
Curvature	01/1985-09/2005	-0.327	0.825	-2.375	1.602	0.910	0.421	-0.079
	01/1995-09/2005	-0.400	0.939	-2.375	1.602	0.895	0.440	-0.035

Note: We present descriptive statistics for the two forecast targets: monthly Personal Income (PI) growth and CPI inflation, and for the yield curve empirical level, slope and curvature, over three different sample periods: full sample ranges from 02/1970 to 09/2005, the 1st out-of-sample evaluation period ranges from 01/1985 to 09/2005, and the 2nd out-of-sample evaluation period ranges from 01/1995 to 09/2005. We define Level as the 10-year yield, Slope as the difference between the 10-year and 3-month yields, and Curvature as the twice the 2-year yield minus the sum of the 3-month and 10-year yields. The last three columns contain sample autocorrelations at displacements of 1, 12, and 30 months.

Table 2. PI Growth Forecast

	Forecasts begin: 01/1985				Forecasts begin: 01/1995			
	h=1	*H*=3	*h*=6	*h*=12	*h*=1	*h*=3	*h*=6	*h*=12
AR	8.382	4.254	3.107	2.505	6.905	3.769	3.045	2.545
APW-OLS	8.339	4.132	2.982	2.427	**6.661**	3.462	2.767	2.390
CF-Mean	8.187	3.955	2.969	2.507	6.899	3.397	2.697	2.311
CF-Median	8.191	3.955	2.967	2.505	6.896	3.393	2.690	2.298
CF-RA (κ=0)	10.123	4.820	3.124	3.070	10.219	3.838	2.766	2.370
CF-RA (κ=1)	9.308	4.362	2.853	2.561	9.578	3.583	2.567	2.203
CF-PC (AIC)	8.900	4.292	2.881	2.954	6.821	3.716	2.747	2.363
CF-PC (BIC)	8.336	4.457	2.619	3.495	6.817	**3.122**	2.737	2.426
CF-PC (1st)	**8.054**	3.642	2.549	1.980	6.817	3.150	**2.355**	**1.839**
CF-PC (2nd)	9.813	6.594	6.089	5.823	8.550	5.972	5.637	5.588
CF-PC (3rd)	9.549	6.421	6.021	6.001	8.215	5.784	5.551	5.331
CF-PC (*k*=2)	8.065	3.608	2.483	1.816	6.843	3.216	2.455	2.000
CF-PC (*k*=3)	8.061	3.619	2.503	**1.780**	6.930	3.203	2.433	2.015
CF-NS-Slope	8.179	3.945	2.957	2.497	6.893	3.392	2.692	2.293
CF-NS-Curvature	8.187	3.956	2.969	2.508	6.898	3.395	2.694	2.307
CF-NS-Level+Slope	8.062	**3.601**	**2.470**	1.845	6.839	3.208	2.447	1.994
CF-NS-All	8.057	3.620	2.514	1.798	6.933	3.195	2.459	2.063
CF-NS-All-Mean	8.184	3.952	2.965	2.503	6.896	3.394	2.694	2.303
CF-NS-All-BMA	8.184	3.951	2.964	2.502	6.896	3.394	2.694	2.303
CF-NS-All-Yang	8.184	3.950	2.961	2.495	6.896	3.394	2.693	2.298
CF-Empirical-Measures	8.067	3.621	2.498	1.812	6.930	3.205	2.447	2.023
CF-NS-Factors	8.089	3.613	2.509	1.905	6.895	3.275	2.511	1.996
CI-Unrestricted	9.456	4.213	2.850	2.211	7.574	3.888	3.212	2.896
CI-PC (AIC)	8.203	4.087	2.846	2.170	6.955	3.842	3.204	2.892
CI-PC (BIC)	8.195	3.944	2.930	2.262	6.894	3.469	3.077	2.905
CI-PC (*k*=1)	8.188	3.956	2.969	2.509	6.893	3.386	2.684	2.295
CI-PC (*k*=2)	8.173	3.932	2.953	2.485	6.972	3.513	2.860	2.516
CI-PC (*k*=3)	8.190	3.952	2.975	2.463	6.995	3.559	2.911	2.616
CI-NS-Level	8.224	4.029	3.057	2.585	6.971	3.507	2.830	2.503
CI-NS-Level+Slope	8.172	3.932	2.954	2.489	6.980	3.528	2.882	2.557
CI-NS-All	8.200	3.963	2.983	2.481	6.997	3.560	2.913	2.612

Note: Forecast target $y_{t+h} \equiv 1200[(1/h)\ln(\mathrm{PI}_{t+h}/\mathrm{PI}_t)]$.

Table 3. CPI Inflation Forecast

	Forecasts begin: 01/1985				Forecasts begin: 01/1995			
	h=1	*H*=3	*h*=6	*h*=12	*h*=1	*h*=3	*h*=6	*h*=12
AR	2.449	1.904	1.646	1.533	2.762	2.181	1.755	1.609
IMA(1,1)-10-year	**2.410**	1.832	1.543	1.448	**2.531**	**1.515**	**1.047**	0.870
CF-Mean	2.820	2.230	2.053	2.031	2.892	2.094	1.895	1.924
CF-Median	2.836	2.252	2.082	2.065	2.889	2.091	1.887	1.909
CF-RA (κ=0)	2.877	1.827	1.395	1.208	3.081	1.942	1.571	1.594
CF-RA (κ=1)	2.645	**1.682**	1.314	1.182	2.939	1.830	1.470	1.477
CF-PC (AIC)	2.576	1.817	1.377	1.199	2.557	1.936	1.557	1.542
CF-PC (BIC)	2.522	1.733	1.341	1.088	2.596	1.568	1.389	1.693
CF-PC (1st)	2.532	1.773	1.412	1.272	2.588	1.569	1.107	0.873
CF-PC (2nd)	4.135	3.598	3.331	3.206	3.578	2.867	2.617	2.482
CF-PC (3rd)	3.881	3.358	3.185	3.095	3.578	2.882	2.612	2.448
CF-PC (*k*=2)	2.550	1.768	1.378	1.231	2.586	1.557	1.090	**0.860**
CF-PC (*k*=3)	2.501	1.689	1.300	1.216	2.608	1.598	1.177	0.992
CF-NS-Slope	2.777	2.176	1.998	1.991	2.883	2.071	1.848	1.851
CF-NS-Curvature	2.825	2.236	2.061	2.039	2.892	2.093	1.893	1.920
CF-NS-Level+Slope	2.552	1.769	1.378	1.233	2.587	1.558	1.091	**0.860**
CF-NS-All	2.494	**1.682**	**1.293**	1.218	2.593	1.582	1.166	0.992
CF-NS-All-Mean	2.806	2.213	2.036	2.019	2.888	2.085	1.878	1.898
CF-NS-All-BMA	2.804	2.211	2.035	2.019	2.888	2.085	1.878	1.898
CF-NS-All-Yang	2.801	2.206	2.032	2.025	2.883	2.078	1.877	1.912
CF-Empirical-Measures	2.503	1.704	1.307	1.209	2.604	1.589	1.166	0.997
CF-NS-Factors	2.542	1.756	1.310	**1.022**	2.618	1.609	1.197	1.000
CI-Unrestricted	2.805	2.109	1.830	1.624	3.069	2.469	2.547	2.629
CI-PC (AIC)	2.667	2.090	1.757	1.616	3.146	2.533	2.592	2.668
CI-PC (BIC)	2.797	2.075	1.744	1.487	3.403	2.638	2.579	2.676
CI-PC (*k*=1)	2.834	2.249	2.078	2.062	2.870	2.059	1.852	1.879
CI-PC (*k*=2)	2.796	2.108	1.798	1.690	3.642	3.023	2.860	2.832
CI-PC (*k*=3)	2.779	2.100	1.792	1.674	3.485	2.880	2.756	2.807
CI-NS-Level	3.009	2.438	2.233	2.146	3.108	2.412	2.306	2.423
CI-NS-Level+Slope	2.784	2.087	1.771	1.676	3.587	2.965	2.815	2.826
CI-NS-All	2.779	2.101	1.796	1.688	3.497	2.877	2.745	2.790

Note: Forecast target $y_{t+h} \equiv 1200[(1/h)\ln(CPI_{t+h}/CPI_t)]$.

Table 4. Forecast Evaluation of IMA(1,1) vs. CF-NS

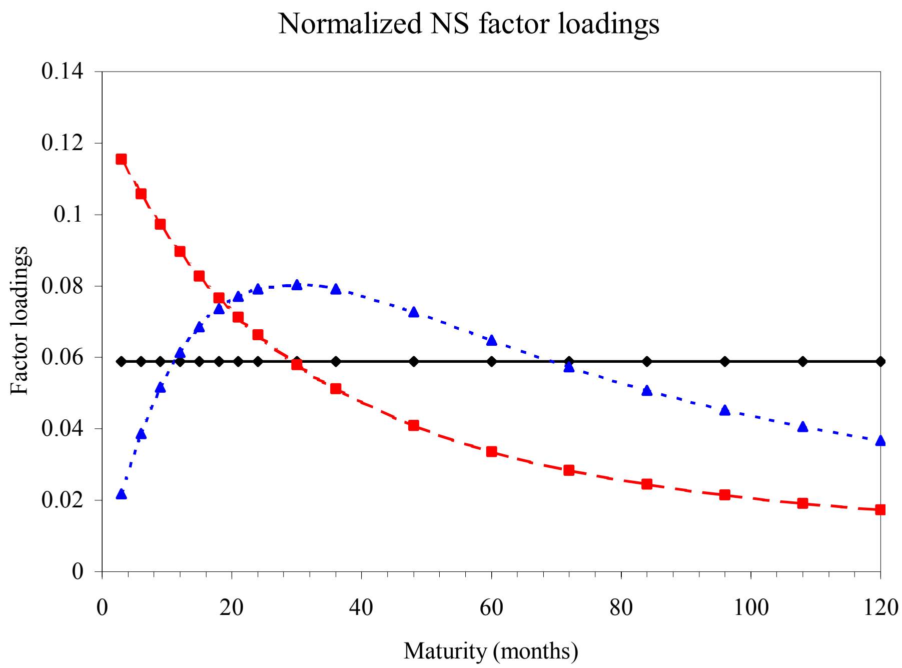
	Forecasts begin: 01/1985			Forecasts begin: 01/1995		
	d_0	d_1	d_2	d_0	d_1	d_2
h=1	1.190	0.521	0.113	1.864	0.100	0.230
	(0.428)	(0.109)	(0.173)	(0.846)	(0.477)	(0.355)
h=3	1.340	0.249	0.326	2.044	-0.050	0.279
	(0.420)	(0.081)	(0.139)	(0.701)	(0.365)	(0.293)
h=6	1.379	0.181	0.363	1.914	0.008	0.252
	(0.365)	(0.068)	(0.119)	(0.609)	(0.328)	(0.238)
h=12	1.602	0.163	0.283	2.246	-0.119	0.223
	(0.327)	(0.087)	(0.122)	(0.510)	(0.244)	(0.163)

Note: We present the estimated regression coefficients (with Newey and West (1987) standard errors in parenthesis) from regressing CPI inflation y_{t+h} on a constant, forecast by IMA(1,1), and the CF-NS forecast:

$$y_{t+h} = d_0 + d_1 \hat{y}_{t+h}^{IMA} + d_2 \hat{y}_{t+h}^{CF-NS} + u_{t+h},$$

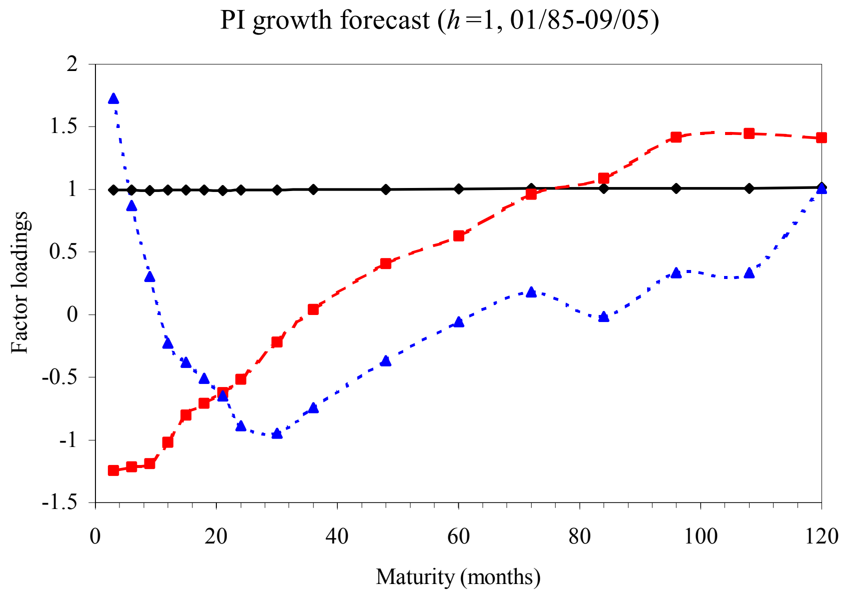
over the two out-of-sample periods at forecast horizon h=1,3,6, and 12. In the first out-of-sample period where forecasts begin 01/1985 the CF-NS forecast used in the regression is from CF-NS-All, while in the second one starts from 01/1995 we use forecast from CF-NS-Level+Slope.

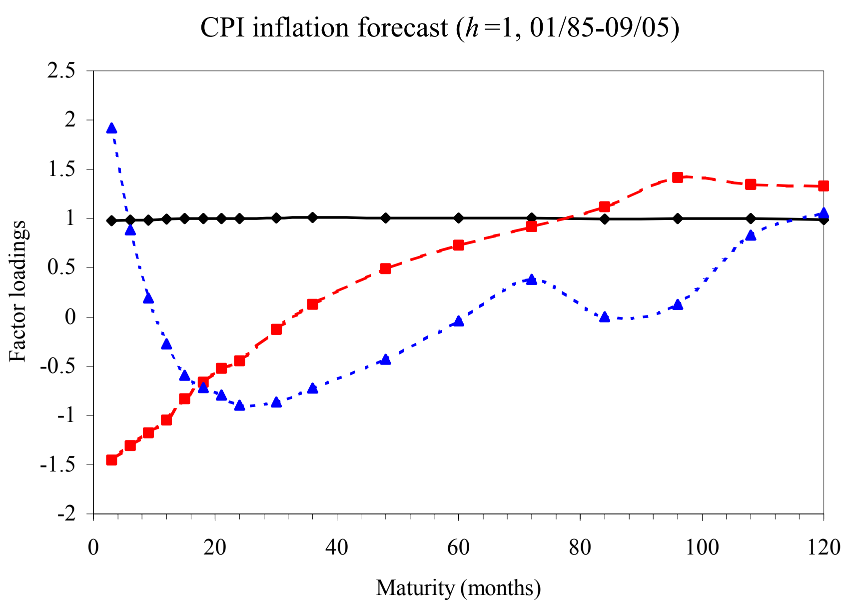
Figure 1. Normalized Nelson-Siegel Exponential Loadings in CF-NS



Note: We present the three normalized Nelson-Siegel (NS) exponential loadings in CF-NS that correspond respectively to the three NS factors. The horizontal axis refers to the 17 individual forecasts that use yields at the 17 maturities (in months). Solid Line denotes the first normalized NS factor loading ($1/N$), Dashed Line denotes the second normalized NS factor loading ($(1-e^{-\lambda\tau})/\lambda\tau$ divided by the sum), and Dotted Line denotes the third normalized NS factor loading ($(1-e^{-\lambda\tau})/\lambda\tau - e^{-\lambda\tau}$ divided by the sum), where τ denotes maturity and λ is fixed at 0.0609.

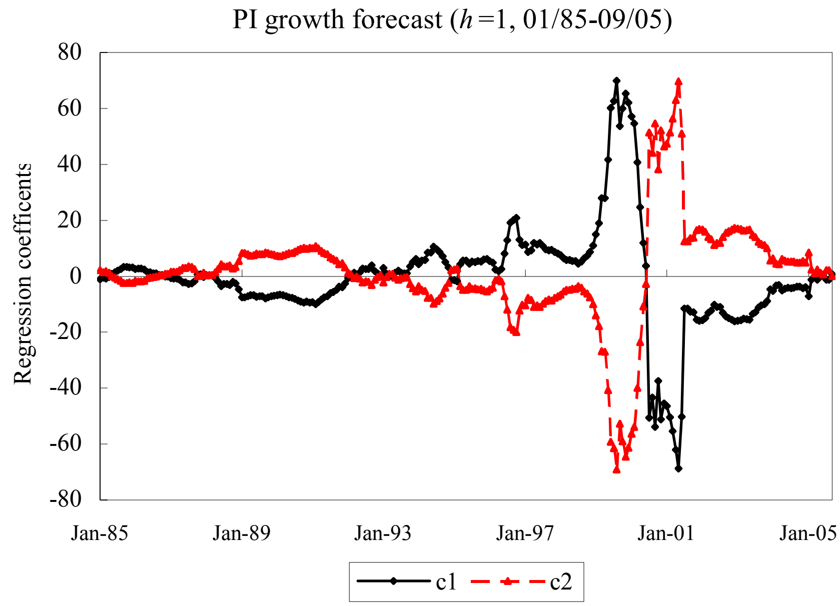
Figure 2. Factor Loadings of the First Three Principal Components in CF-PC

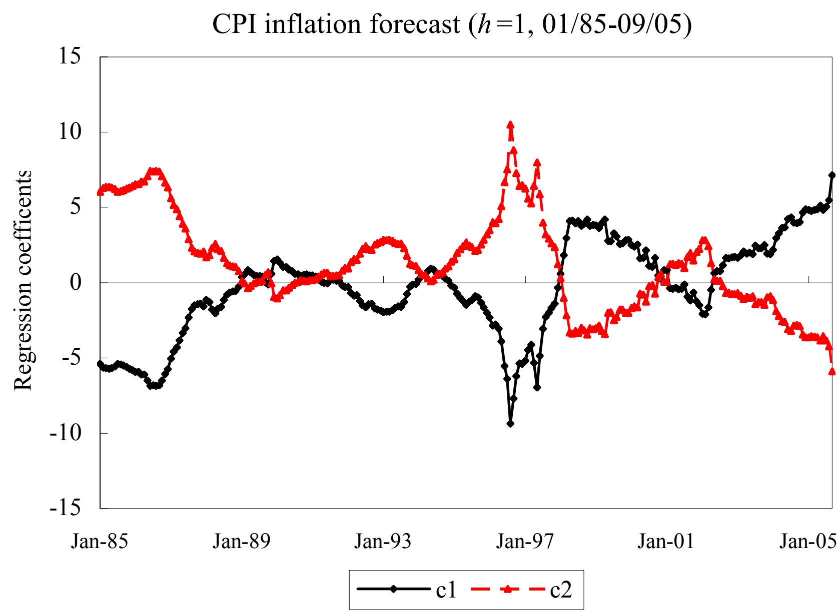




Note: We present the factor loadings of the first three principal components in CF-PC methods averaged over the entire first out-of-sample period (01/1985-09/2005), for both PI growth and CPI inflation forecasting. The horizontal axis refers to the 17 individual forecasts that use yields at the 17 maturities (in months). Solid Line denotes the loading of the first PC, Dashed Line denotes the loading of the second PC, and Dotted Line denotes the loading of the third PC.

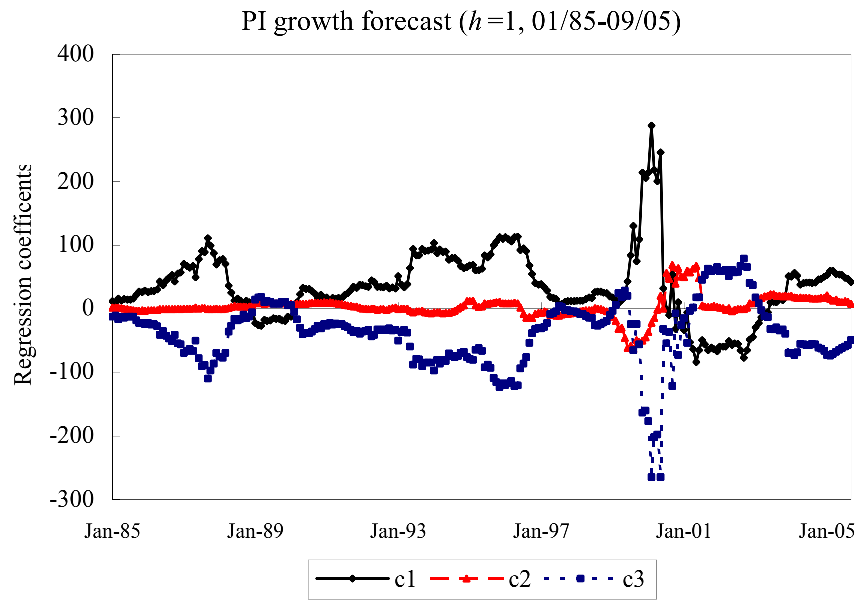
Figure 3. Estimated Regression Coefficients in CF-NS-Level+Slope

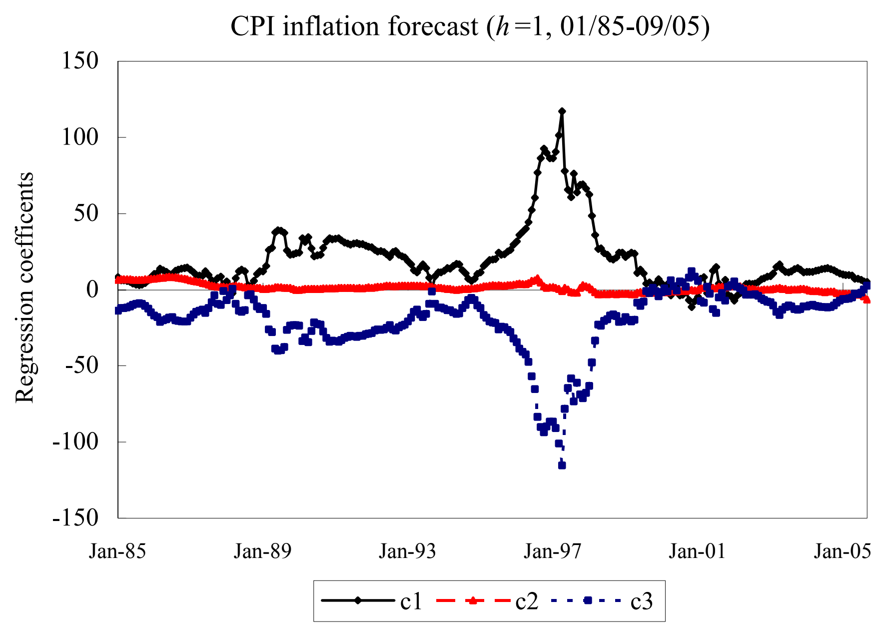




Note: We present the estimated regression coefficients in CF-NS-Level+Slope over the entire first out-of-sample period (01/1985-09/2005), for both PI growth and CPI inflation forecasting. Solid Line (c1) denotes the coefficient for the first combined forecast z_1, and Dashed Line (c2) denotes the coefficient for the second combined forecast z_2.

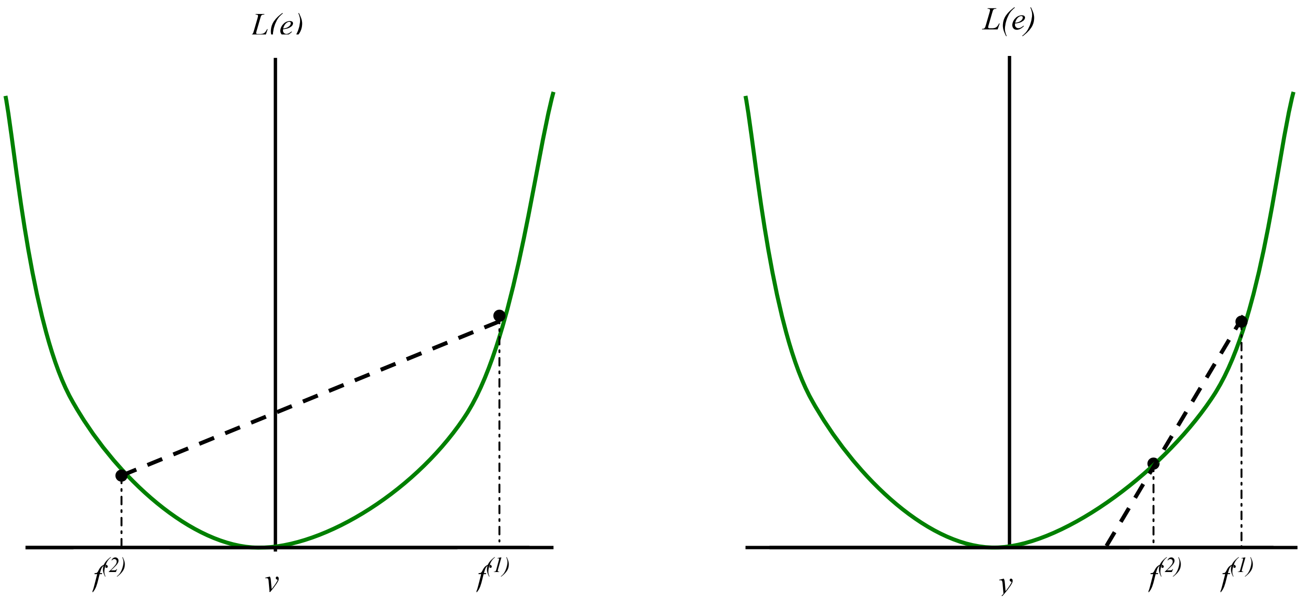
Figure 4. Estimated Regression Coefficients in CF-NS-All





Note: We present the estimated regression coefficients in CF-NS-All over the entire first out-of-sample period (01/1985-09/2005), for both PI growth and CPI inflation forecasting. Solid Line (c1) denotes the coefficient for the first combined forecast z_1, Dashed Line (c2) denotes the coefficient for the second combined forecast z_2, and Dotted Line (c3) denotes the coefficient for the third combined forecast z_3.

Figure 5. Optimal Forecast Combination Depending on ρ



(a)
$\rho < 0$
$0 < k < 1$

(b)
$\rho > 0$
$k < 0$