MACQUARIE UNIVERSITY
FACULTY OF
BUSINESS AND ECONOMICS

Centre for Financial Risk

Yield curve dynamics in low interest rate environments –

The unbeatable random walk?

Dennis Wellmann, Stefan Trück

Working Paper 15-02

The Centre for Financial Risk brings together Macquarie University's researchers on uncertainty in capital markets to investigate the nature and management of financial risks faced by institutions and households.

Research conducted by members of the centre straddles international and domestic issues relevant to all levels of the economy, including regulation, banking, insurance, superannuation and the wider corporate sector, along with utility and energy providers, governments and individuals.

The nature and management of financial risks across these diverse sectors are analysed and investigated by a team of leading researchers with expertise in economics, econometrics and innovative modelling approaches.

Yield curve dynamics in low interest rate environments - the unbeatable random walk?

Stefan Trück[a*], Dennis Wellmann[a]

[a]Faculty of Business and Economics, Macquarie University, Sydney, NSW 2109, Australia

This version: January 2015

Abstract

We investigate the forecasting performance of popular dynamic factor models of the yield curve after the global financial crisis (GFC). This time period is characterized by an unprecedented low and non-volatile interest rate environment in most major economies. We focus on the dynamic Nelson-Siegel model and regressions on principal components and use a dataset of monthly US treasury bond yields to show that subsequent to the GFC both models are significantly outperformed by the random walk no-change forecast. Especially for short and medium term yields the random walk is up to ten times more accurate. Interestingly, these results are not picked up by traditional global forecast evaluation metrics. We show that combining forecasts mitigates the model uncertainty and improves the disappointing forecasting accuracy especially after the GFC.

Key words: Term structure of interest rates, yield curve forecasting, Nelson-Siegel model, dynamic factor models

JEL: C32, C53, E43, E47

*Corresponding author. Contact: Stefan.Trueck@mq.edu.au ; +61 2 98508483.

1 Introduction

The global financial crisis (GFC) in 2007-2009 has caused major eruptions in bond and interest rate markets, rendering many traditional yield and bond pricing models useless (Bianchetti, 2010; Walker and McCormick, 2014). The GFC has also led to an unprecedented, prolonged period of low interest rates in several advanced economies subsequent to the crisis. The US are a prime example for this development. Following the Fed's expansive monetary policy during the GFC, US short and medium yields have been more or less flat since 2009. We show how this unique interest rate environment severely challenges the forecasting accuracy of popular dynamic factor yield curve forecasting models for short and medium yields. We also show that the poor forecasting performance in this time period is not reflected in traditional, global forecast evaluation metrics such as the root mean squared error (RMSE) that are typically applied to measure the forecasting performance, see, e.g. Diebold and Li (2006); Carriero et al. (2012). Hence this outcome may not be perceived in future forecasting studies and potentially distort results and interpretations. Finally, we suggest forecast combination strategies as a mitigating measure to improve the forecasting accuracy.

Forecasting the yield curve has been recognized in the literature as a challenging task for decades. Despite major advances in yield curve modelling and forecasting (Vasicek, 1977; Cox et al., 1985; Nelson and Siegel, 1987; Dai and Singleton, 2000; Diebold and Li, 2006; Exterkate et al., 2013), the high persistence of yields makes it typically hard for any model to outperform a simple random walk no-change forecast (Ang and Piazzesi, 2003; Moench, 2008; Carriero et al., 2012; Xiang and Zhu, 2013). In this study we illustrate that the current interest rate environment even further aggravates this challenge. After the GFC, US short and medium yield forecasts of popular dynamic factor models not only fail to beat the random walk but are completely outperformed in relative terms, with a random walk model being up to ten times more accurate. Surprisingly, this outcome has not been thoroughly investigated and documented yet in previous research.

To investigate this effect we study a dataset of monthly US Treasury bonds zero-coupon yields for the time period from 1995:01 to 2013:12, focusing on the class of dynamic factor models. While they may lack the theoretical fundament of no-arbitrage models, they promise to deliver the most accurate forecasting results as suggested by Duffee (2002, 2011); Chen and Niu (2014)), just to name a few. They are also the class predominantly used in recent studies focusing on predicting the term structure of interest rates. We focus on different variations of the Nelson-Siegel model, which imposes a parametric structure on factor loadings, as well as regressions on principal components, which extract factors and factor loadings directly from the data. In our analysis we benchmark the forecasting performance of these models against a random walk no-change forecast. We also include a simple autoregressive (AR(1)) model as an additional benchmark, since this approach has been reported to forecast the yield curve surprisingly well (Diebold and Li, 2006; Pooter et al., 2010). The forecasting accuracy is measured with the commonly used root mean squared error (RMSE) and Diebold-Mariano statistics (DM).

We start our analysis with investigating the forecasting accuracy for the entire forecasting period from 2004:01 to 2013:12 and find results similar to previous comprehensive US yield curve forecasting studies, see, e.g. Pooter et al. (2010) or Yu and Zivot (2011). The selected factor models perform relatively well for short maturities and long forecast horizons, but all models fail to consistently beat the random walk. However, the sub-sample analysis reveals that subsequent to the GFC (2009:01-2013:12) the forecasting accuracy for short and medium yields worsens dramatically relative to the random walk. For nearly all maturities below five years the random walk is multiple times more accurate across all forecasting horizons. For six and twelve-months ahead forecasts the random walk is even up to ten times more accurate. Similar results are also found when comparing the performance of the applied dynamic factor relative to an AR(1) model. In addition, Diebold-Mariano statistics show, that the considered models provide forecasts that are significantly worse than those of a random walk. In other words, since the end of the GFC the random walk and a simple AR(1) process significantly outperform popular yield curve forecasting models in predicting short- and medium term yields. While the performance of forecasting models naturally varies over time, these results are still striking over such an extended period. We argue that since the applied dynamic models are typically calibrated over a period that also includes significant changes in interest rates as well as in the term structure of the yield curve, these models seem to be outperformed by a no-change random walk forecast in a low yield environment with hardly any fluctutions for the observed interest rates. Moreover, the models were also estimated during periods when interest rates were significantly higher than during the post GFC period such that forecasts created by the applied models may not only overstate the dynamics of the interest rate term structure but also interest rate levels.

Interestingly, this performance is not picked up by commonly used global forecast evaluation metrics when the entire out-of-sample period is considered such that the results are not reflected in the full sample RMSE's. This is because the forecasting errors during the critical period become relatively small in absolute terms, especially for short and medium yields and thus contribute relatively little to the global average. This highlights one of the most important points of this paper: investigating the global (or average) absolute forecasting performance may hide important information about the relative forecasting performance over time.

A natural question to ask is how to approach the unique yield curve dynamics subsequent to the GFC in future forecasting exercises. Different approaches have been developed to account for structural instability. Ang and Bekaert (2002) or Xiang and Zhu (2013), for example, suggest to apply regime-switching models that may capture the different interest rate environment. Exterkate et al. (2013) have also shown that including macroeconomic factors may improve the forecasting performance especially in volatile time periods, while gains in the forecasting performance are clearly less significant during when volatility is low. We suggest forecast combination techniques (Timmermann (2006); Guidolin and Timmermann (2009); Pooter et al. (2010)) as a possible strategy to mitigate the model uncertainty and improve the disappointing forecasting accuracy, especially for the crucial time period after the GFC. We find that simply combining all applied factor models already significantly improves the forecasting accuracy compared to the individual models, albeit this strategy is still outperformed by a random walk. We also combine two diametrically biased variations of Nelson-Siegel and principal component model with the AR(1) model and find that this strategy is able to further improve the poor forecasting performance for shorter maturities after the GFC. This strategy is also able to beat the random walk for longer forecasting horizons. Our results also indicate that performance weighted forecast combination schemes generally lead to more accurate forecasts than the equally weighted performance schemes.

We contribute to the literature on yield curve forecasting in several dimensions. To begin with, this is the first paper to systematically document and explain the poor forecasting performance for medium and short term yields associated with the popular class of dynamic factor yield curve models in the current low interest rate environment. While we focus on the most popular, basic variations of the models, further research may be required to examine how alternative and more complex dynamic factor models perform in this time period.

Second, we show how sensitive the forecasting performance is to the choice of the evaluation metrics. It is still common to select the model with the best global forecasting performance, which in practice amounts to selecting the model that forecasts best on average over the entire out-of-sample period. However, in the presence of time-varying yield curve dynamics, averaging the results over time will result in a significant loss of information.

Third, we provide further evidence that combining different models can significantly improve the forecasting accuracy, especially in periods where individual models perform poorly. While the forecasting accuracy of the selected models varies heavily over time, forecast combinations are less affected by structural instability than either of the individual models.

Finally, our results illustrate how important it is to closely examine the dynamic behavior of the yield curve and to perform a thorough sub-sample analysis and apply dynamic forecast evaluation measures to reveal the true forecasting performance in future yield curve prediction exercises.

The remainder of the paper is organized as follows: Section 2 provides a review of the relevant yield curve forecasting literature. Section 3 reports descriptive statistics and illustrates the dynamic behavior of yields during the considered sample period. In Section 4 we introduce the selected models, while Section 5 provides out-of-sample forecasting results and several robustness checks. In Section 6 we apply different forecast combination strategies and examine whether results can be improved in comparison to using individual models only. Finally, Section 7 concludes and provides suggestion for future work in the area of research.

2 Related Literature

The numerous term structure models can typically be divided into two streams of literature, see, for example Chen and Niu (2014). The first stream consists of models deriving the term structure based on the short rate, by eliminating arbitrage possibilities between current and future interest rates under various assumptions about the risk premium. Building on the work of Vasicek (1977) and Cox et al. (1981), seminal contributions to the development of these no-arbitrage and affine equilibrium models include Hull and White (1990); Duffie and Kan (1996) and Dai and Singleton (2000). More recent contributions to this stream of literature also relate the short rate to macroeconomic variables (Ang and Piazzesi, 2003; Dewachter and Lyrio, 2006; Rudebusch and Wu, 2008; Moench, 2008). Unfortunately, no-arbitrage and affine-equilibrium models often exhibited poor empirical forecasting performance as pointed out by Duffee (2002, 2011).

The second stream of literature consists of reduced-form models based on more data-driven statistical approaches. This stream has evolved from univariate to multivariate time series models to the class of empirical factor models predominantly used today. Popular univariate models are, for example, the slope regression model, the Fama-Bliss forward rate regression model (Fama and Bliss, 1987) or simple autoregressive processes. The multivariate class includes in particular vector autoregressive (VAR) models and error correction models (ECMs). Different to the univariate models, these

models are also able to utilize the cross-sectional dependence structure and cointegration of observed yields at different maturities.
In this study we mainly focus on the class of empirical dynamic factor models that recently have been extensively applied to the modeling and prediction of the yield curve (Christensen et al., 2011; Favero et al., 2012; Exterkate et al., 2013; Xiang and Zhu, 2013). Dynamic factor models allow to model and forecast the term structure based on low-dimensional, latent factors which are extracted from the entire yield curve while retaining the dependence structure of different maturities. The latent factors are usually either estimated by imposing a parametric structure on the factor loadings or extracted directly from the term structure, e.g., by means of a principal component analysis (PCA). While these models may lack the theoretical foundation of the first stream, the empirical literature suggests that they may provide more accurate forecasts of the yields (Duffee, 2002; Pooter et al., 2010).
Most of the parametric factor models build on the ground-breaking work of Nelson and Siegel (1987) and Diebold and Li (2006). Nelson and Siegel (1987) introduced a parsimonious three-factor model to fit the term structure by using flexible, smooth parametric functions. They demonstrate that their model is capable of capturing most of the typically observed shapes assumed by the yield curve over time. Among the various extensions that have been proposed to incorporate additional flexibility, the most popular is probably the Svensson (1994) four-factor model. Both, the Nelson-Siegel as well as the Svensson model are heavily used by market practitioners and central banks to construct zero-coupon yield curves, see, for example, Gürkaynak et al. (2007); Coroneo et al. (2011)).
The initial Nelson-Siegel model only estimated the yield curve at certain point in time. Diebold and Li (2006) have extended Nelson-Siegel's initial approach into a dynamic framework and applied it successfully to forecast the term structure of US yields. Their parsimonious three-factor model performs surprisingly well, particularly at long horizons, while the three latent factors in the model can be reinterpreted as the level, slope and curvature of the yield curve.
Since the seminal study by Diebold and Li (2006), the literature on fore-

casting yield curves has grown significantly and in particular the dynamic Nelson-Siegel model has been extended numerous times. Diebold et al. (2006) integrate the initial Diebold and Li (2006) two-step forecasting approach into a single dynamic factor model by specifying the Nelson-Siegel weights as an unobserved vector autoregressive process. Diebold et al. (2008) further extend the initial dynamic Nelson-Siegel model to a global context in which modeling a large set of yield curves allows for global and country specific factors. Christensen et al. (2011) develop an arbitrage-free version of this model, while Yu and Zivot (2011) include the evaluation of a state-space approach and nine different ratings for corporate bonds. Hautsch and Yang (2012) allow for stochastic volatility of the estimated yield factors, while Xiang and Zhu (2013) develop a regime-switching Nelson-Siegel model. Most recently, Laurini and Hotta (2014) and Chen and Niu (2014) integrate Bayesian estimation methods and adaptive forecasting techniques into the dynamic factor framework.

An alternative, non-parametric forecasting approach is to apply PCA to extract the factors directly from the entire term structure. PCA works best with correlated time series (Duffee (2012) and is therefore a natural and popular choice to reduce the dimensions of highly correlated yield curve datasets. A small number of orthogonal and uncorrelated factors or principal components can usually account for a high fraction of variability in relatively high-dimensional datasets. Following Litterman and Scheinkman (1991), several studies apply PCA and find that the variation in interest rates can already be explained by the first three principal components, see, e.g., Bikbov and Chernov (2010); Leite et al. (2010). These three common factors also have an intuitive interpretation as level, slope and curvature based on their effect on the yield curve and can be successfully applied in forecasting exercises. Reisman and Zohar (2004), for example, use their forecasting results in bond portfolio selection and suggest that frequent rebalancing leads to substantially higher returns. Blaskowitz and Herwartz (2009) apply PCA to the prediction of the term structure of Euribor swap rates.

While dynamic factor models are the most promising class of yield curve forecasting models the near unit root behaviour of the yields makes it hard for any model to consistently outperform the random walk. Still, many studies report superior forecasting results for particular datasets, however, as Pooter et al. (2010) show in an extensive forecasting study of US yields, no model clearly performs well across all maturities or different sample periods. Moreover, the forecasting ability of individual models considerably varies over time.

Recent studies have shown that combining the forecasts of different models may mitigate this model uncertainty (Guidolin and Timmermann, 2009). A different approach may also be to include macroeconomic variables into the forecasting procedure. Amongst others, Koopman and van der Wel (2013) and Exterkate et al. (2013) have demonstrated that including macroeconomic variables can significantly improve the forecasting performance for yield term structures, especially during periods of poor forecasting accuracy.

Overall, despite recent advances, forecasting the yield curve remains a challenging task. In this study we show that forecasting short and medium yields becomes even more arduous in the current low-interest rate environment after the GFC.

3 Data

For the analysis, we use the end-of-month zero-coupon rates of US Treasury bonds obtained from Bloomberg for the time period from January 2000 to December 2013. Selecting US yields is an obvious choice as they have predominantly been used in the literature due to their supreme data quality and availability. The US are also a prime example for an extended period of low and non-volatile interest rates after the GFC. Bloomberg has the advantage of providing up to date yields and thus including the time period after the GFC. Using monthly frequency (n=228) we construct the term structure with 12 maturities ranging from 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108 to 120 months.

Table 1 provides the descriptive statistics of the considered dataset. The reported characteristics are in line with the stylized facts commonly found in yield curve data, see e.g Diebold and Li (2006); Pooter et al. (2010) or

Maturity (months)	Mean	St Dev	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	$\hat{\alpha}(2)$	$\hat{\alpha}(12)$	ADF
3	2.85	2.26	0.02	6.47	0.99	0.74	0.25	-0.27	-0.05	-1.25
6	3.00	2.32	0.04	6.74	0.99	0.74	0.25	-0.35	-0.05	-1.36
12	3.12	2.32	0.11	6.88	0.99	0.75	0.27	-0.27	-0.10	-1.29
24	3.33	2.22	0.22	7.48	0.98	0.76	0.33	-0.18	-0.07	-1.28
36	3.55	2.11	0.28	7.59	0.98	0.77	0.37	-0.16	-0.06	-1.42
48	3.77	2.00	0.44	7.68	0.98	0.77	0.40	-0.12	-0.06	-1.66
60	3.95	1.85	0.62	7.72	0.97	0.76	0.40	-0.10	-0.05	-1.80
72	4.12	1.76	0.79	7.79	0.97	0.75	0.41	-0.10	-0.05	-1.90
84	4.31	1.67	0.97	7.86	0.97	0.74	0.42	-0.10	-0.05	-2.02
96	4.44	1.58	1.17	7.87	0.97	0.74	0.42	-0.10	-0.06	-2.07
108	4.54	1.50	1.37	7.89	0.97	0.73	0.41	-0.10	-0.08	-1.89
120	4.60	1.42	1.60	7.90	0.97	0.70	0.39	-0.08	-0.07	-2.30

Table 1. Descriptive Statistics for the term structure of US yields for the time period from 2000:01 to 2013:12. For each maturity we report (from left to right) mean, standard deviation, minimum, maximum, autocorrelations at displacements of 1, 12, and 30 months, partial autocorrelations at displacements of 2 and 12 months and augmented Dickey-Fuller (ADF) test-statistics. For the ADF, the critical values for a rejection of the unit root hypothesis are 3.45 at the 1% level (indicated by ***), 2.87 at the 5% level (**) and 2.57 at the 10% level (*). SIC is applied to determine the lag length.

Koopman and van der Wel (2011). The average yield curve during the sample period is upward sloping and concave, volatility is decreasing with maturity and autocorrelations are very close to unity. The ADF statistics confirm that yields are indeed all but non-stationary. The partial autocorrelation function suggests that autoregressive processes of limited lag order may fit the data well. Correlations between yields of different maturities are not reported here but are typically high, especially for adjacent maturities.

In Figure 1, we plot the dynamic behavior for yields of selected maturities. The plot confirms that the yield curve is mostly upward sloping with only two short periods of inverted yield curves preceding the two recessions (March - November 2001 and December 2007 - June 2009) after the bursting of the dotcom bubble and the GFC period. These periods also reveal that short and long maturities react quite differently to economic shocks as both recessions are characterized by a sharp decline in short yields and, thus, an increase in the spread between short and long yields. The term spread is generally known to remain rather large for quite some time after recessions. Nevertheless, the behavior of short and medium interest rates, e.g., three-months to 36-months, after the GFC is startling. Following the Federal

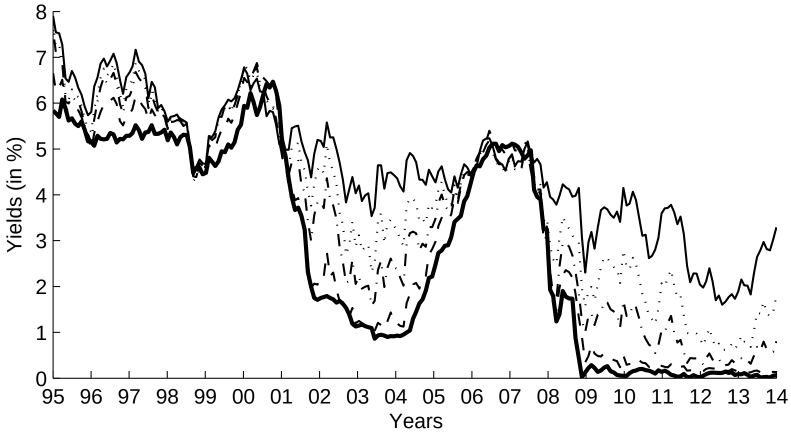


Figure 1. Time series of selected US yields. We plot the three-months (bold —), twelve-months (– –), 36-months (– ·), 60-months (· · ·) and 120-months maturities (—) for the time period 2000:01 – 2013:12.

Reserve Bank's unprecedented expansive monetary policy in response to the crisis, short yields remain flat and non-volatile for more than five years up until now. Medium-term yields behave similar reflecting the Fed's strong commitment to maintaining the expansionary policy as long as required for economic recovery.[2] Assisted by several programs of 'quantitative easing'[3] this has led to an unprecedented, prolonged period of low, non-volatile short and medium yields. This unique interest rate environment is expected to favor a random walk no-change forecast and we expect that it will pose a peculiar challenge for the forecasting models introduced in the subsequent sections.

4 Models

In Section 2 we have described the numerous empirical factor models that have been developed to model and forecast the yield curve in recent decades.

[2]See, for example, chairman Bernanke's famous quote "The Federal Reserve has done, and will continue to do, everything possible within the limits of its authority to assist in restoring our nation to financial stability", when speaking at the National Press Club in 2009.

[3]The acquisition of financial assets from commercial banks to lower longer yields while simultaneously increasing the monetary base.

To keep the number of models manageable, we focus on a representative subset of basic models which are commonly used in the academic literature and by practitioners. In particular we include one model imposing a parametric structure on factor loadings, the dynamic Nelson-Siegel model, and PCA as a model that extracts the loadings and factors directly from the observed term structure. For both models we apply AR(1) and VAR(1) factor dynamics. While the jointly specified VAR(1) process has the advantage of capturing the interdependence between the derived factors, both approaches have been reported to work well in forecasting exercises (Diebold and Li, 2006; Pooter et al., 2010).

Furthermore, we include an AR(1) model directly applied on yield levels. AR(1) models can be considered as simple workhorse models and have been reported to fit and forecast the yield levels quite well. The models applied in our empirical analysis are specified as follows:

Dynamic Nelson-Siegel Model

The Nelson and Siegel (1987) model is a parsimonious, parametric three-factor model using curve using flexible, smooth Laguerre functions to estimate the yield curve. Based on the three parametric loadings, $[1, (1 - e^{-\lambda_t \tau})/\lambda_t \tau, ((1 - e^{-\lambda_t \tau})/\lambda_t \tau) - e^{-\lambda_t \tau}]$, the yield y_t for maturity τ for the dynamic Nelson-Siegel model developed by Diebold and Li (2006) is modeled as

$$y_{t,\tau} = \beta_{1,t} + \beta_{2,t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3,t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right), \tag{1}$$

where $\beta_{1...3,t}$ denote the three latent factors, and the parameter λ controls the exponential decay rate of the second and third loading. In line with Diebold and Li (2006), Diebold et al. (2008) and Chen and Gwati (2013) we fix λ at 0.0609.

To forecast the term structure, we follow Diebold and Li's (2006) two-step approach.[4] First, the Nelson-Siegel factors $\beta_{1...3}$ are estimated for the in-sample period applying ordinary least squares. Then the factors are forecasted as au-

[4] Note that we refrain from using the one-step state-space approach. Pooter et al. (2010), for example, report no substantial gain in forecasting accuracy across maturities and horizons.

toregressive processes, i.e. for the AR(1) approach each $\hat{\beta}_{k,t+h/t}$ is forecasted as

$$\hat{\beta}_{k,t+h/t} = \hat{c}_{k,h} + \hat{\phi}_{k,h}\hat{\beta}_{k,t}, \tag{2}$$

while for the VAR(1) factor dynamics, each $\hat{\beta}_{k,t+h/t}$ is forecasted as

$$\hat{\beta}_{k,t+h/t} = \hat{c}_{k,h} + \hat{\Phi}_{k,h}\hat{\beta}_{k,t}, \tag{3}$$

such that each individual yield forecast for maturity τ is given by

$$\hat{y}_{t+h/t,\tau} = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \hat{\beta}_{3,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right). \tag{4}$$

In the following we will denote the two approaches by **NSAR** and **NSVAR**.

Regression on principal components

For the PCA approach, each yield is given by the following dynamic latent factor model:

$$y_{t,\tau} = \gamma_{1,\tau}\beta_{1,t} + ... + \gamma_{K,\tau}\beta_{K,t} + \epsilon_{t,\tau}, \tag{5}$$

where the $\gamma_{K,\tau}$ describe the K factor loadings and $\beta_{K,t}$ represent K vectors of latent factors.[5] The factors and loadings are estimated with a principal component analysis on the full set of yields for every forecasting iteration. Note that we use standardized yields with zero mean and unit variance for the PCA.

To derive the loadings $\gamma_{K,\tau}$ and latent factors $\beta_{K,t}$, a PCA seeks an orthogonal matrix Γ which yields a linear transformation $\Gamma Y = B$ of the T x N dimensional matrix of standardized yields Y and K x N-dimensional matrix B of latent factors $\beta_{K,t}$ such that the maximum variance is extracted from the variables. The matrix Γ is constructed using an eigenvector decomposition. Let Σ denote the T x T covariance matrix of Y. This covariance matrix can be decomposed as

$$\Sigma = \Gamma\Lambda\Gamma', \tag{6}$$

[5]Please note that the terms factor and principal component are used interchangeably throughout this analysis.

where the diagonal elements of $\Lambda = diag(\lambda_1, ..., \lambda_K)$ are the eigenvalues and the columns of Γ are the eigenvectors. The eigenvectors are arranged in decreasing order of the eigenvalues and the first K eigenvectors of Γ denote the factor loadings $[\gamma_1, ..., \gamma_K]$. The K latent factors $[\beta_1, ..., \beta_K]$ are then constructed by $\beta_{k,t} = \gamma_k' Y_t$. Hereby, Y_t is a T-dimensional vector of the term structure of interest rates at time t.
Applying a PCA to extract the latent factors allows for a data-driven selection of the number of K factors. We decide to use the first three latent factors in line with previous research. Typically, the first three principal components are already sufficient to explain a high fraction of the variance in yields (Litterman and Scheinkman, 1991; Bikbov and Chernov, 2010). We find that for the applied dataset, the first three components explain more than 99% in the variation of the term structure. We apply the two-step forecasting procedure outlined above, forecasting the components $\hat{\beta}_{k,i}$ as AR(1) and VAR(1) processes. Thus, h-step ahead yield forecasts for maturity τ are then given by

$$\hat{y}_{t+h/t,\tau} = \gamma_{1,\tau,t}\hat{\beta}_{1,t+h/t} + \gamma_{2,\tau,t}\hat{\beta}_{2,t+h/t} + \gamma_{3,\tau,t}\hat{\beta}_{3,t+h/t}. \tag{7}$$

In the following, we will refer to these models as **PCAAR** and **PCAVAR**.

Autoregressive (AR(1) model on yield levels
We also apply AR(1) models to individual yields of maturity τ directly, determining h-step ahead forecasts as

$$\hat{y}_{t+h/t,\tau} = \hat{c}_{\tau,h} + \hat{\phi}_h S_{t,\tau}, \tag{8}$$

where $\hat{c}_k$ and $\hat{\phi}_k$ are obtained by regressing $\hat{s}_{t,\tau}$ on an intercept and $\hat{y}_{t-h,\tau}$. We denote this model as **AR1**.

Random Walk
As the main benchmark model throughout the forecasting exercise we use a random walk model. In this model any h-step ahead forecast is simply equal

to the value observed at time t. Hence the forecast is always no change and given as

$$\hat{y}_{t+h/t,\tau} = y_{t,\tau}. \tag{9}$$

We denote the random walk benchmark model as **RW**.

5 Out-of Sample Forecasting Results

5.1 Forecasting Framework and Evaluation

In the following we thoroughly investigate the performance of the applied econometric models in forecasting the US yield curve against a random walk benchmark. For the forecasting exercise, the sample of size N is divided into an in-sample period of length R and an out-of-sample period of length P. We use an initial in-sample period from 1995:1 to 2003:12 to forecast the period from 2004:1 to 2013:12. Thus, the in-sample period includes the bursting of the dotcom bubble and the subsequent recession and recovery, while the out-of-sample period includes the GFC as well as pre- and post crisis periods. The considered sample period allows us to have enough observations to estimate the parameters of the models with sufficient accuracy and still evaluate the forecasting performance over sufficiently long (sub-)periods with different yield curve dynamics.

We forecast recursively such that in each time step the in-sample period is extended by one observation to calculate the forecasts for t+h. In particular, we create one-month ($h = 1$), six-month ($h = 6$) and twelve-month ($h = 12$) ahead forecasts whereas all models are forecast iteratively.[6]

To assess the full sample forecasting accuracy we first report the commonly used root mean squared error (RMSE) and Diebold-Mariano (DM) statistic. The RMSE is a measure of global forecasting performance and summarizes

[6]It is still being debated whether iterated or direct forecasts are more accurate. Carriero et al. (2012) for example find that the iterated approach produces more accurate forecasts in yield curve forecasting. Comparing both approaches we also find better results for the iterated approach and henceforth apply it throughout the analysis.

the forecasting errors over the entire forecasting period. For each considered model m, maturity τ and forecasting horizon h the RMSE for the forecasting period P is calculated as

$$RMSE^m_{\tau,h} = \sqrt{\frac{1}{P}\sum_{t=1}^{P}(\hat{y}^m_{t+h/h,\tau} - y^m_{t+h,\tau})^2}. \tag{10}$$

The lower the RMSE, the more accurate the forecast. However, a smaller RMSE in a particular sample of forecasts does not necessarily mean that the corresponding model is truly better in population. Diebold and Mariano (1995) address this concern and propose a test to assess the statistical significance of predictive superiority. The DM-test statistic is calculated as

$$DM^m_{\tau,h} = \frac{\bar{d}}{\sqrt{\widehat{LRV}_{\bar{d}}/P}}, \tag{11}$$

where $\bar{d}$ is the average difference d between the loss functions[7] of two competing forecasts given as

$$\bar{d} = \frac{1}{P}\sum_{t=1}^{P} d_t. \tag{12}$$

$\widehat{LRV}_{\bar{d}}$ is a HAC estimator of the asymptotic (long-run) variance of $\sqrt{P}/\bar{d}$ given by

$$\widehat{LRV}_{\bar{d}} = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j, \tag{13}$$

where $\gamma_0 = var(\bar{d})$ and $\gamma_j = cov(d_t, d_{t-j})$. The null hypothesis is equal predictive accuracy of the considered models. Note that we will conduct two-sided tests, since we are interested in both, statistically significant superior and inferior forecasting performance of the selected models against a random walk benchmark.

[7]We also apply the commonly used quadratic loss functions. However, theoretically Diebold and Mariano (1995) do not limit the loss functions that could be used.

5.2 Forecasting Results

Table 2 presents the forecasting results for the out-of-sample forecasting period from 2004:01 up to 2013:12. In the first line we report the RMSE of the random walk expressed in basis points. We then report the RMSEs of all models relative to the random walk. Hence, numbers smaller than one (**reported in bold**) indicate that a model outperforms the random walk. The significantly better forecasting performance of a model against the random walk benchmark based on conducted DM-tests[8] is indicated by ("), while we indicate the significantly inferior performance of a model against the random walk by (∗).

We find roughly similar outcomes to other comprehensive forecasting studies (Pooter et al., 2010; Yu and Zivot, 2011). In absolute terms the RMSEs are generally smaller for longer term maturities and the forecasting performance worsens with longer forecasting horizons. In relative terms, the applied factor models perform relatively well for the shorter maturities. Nevertheless, all models fail to consistently beat the random walk - not a single model clearly performs well across all maturities and forecast horizons. The Diebold-Mariano statistics confirm that no model is able to significantly outperform the random walk at any maturity.

Given the unique interest rate environment after the GFC, the superior forecasting performance for short and medium yields comes as a surprise. The relatively flat short and medium yields clearly favour the random walk no-change forecast, thus we would have expected the factor models to underperform the random walk. After all half the period after the GFC makes up half of the forecasting period

Comparing the different models, the AR(1) process performs surprisingly well and is on par with the factor models for most maturities and forecast horizons. Noteworthy is also the rather disappointing forecasting performance of the initial dynamic Nelson-Siegel model with AR(1) factor dynamics. Similar disappointing results for the dynamic Nelson-Siegel model have also been

[8]Detailed results and test statistics for the conducted DM-tests are reported in Appendix A.

reported by Pooter et al. (2010) and Moench (2008) who suggest that the reported success of the initial Diebold and Li (2006) model for predicting yield curve dynamics may be attributed to the choice of the forecasting period. In general, capturing the factor dependence structure with VAR(1) factor dynamics seems to lead to slightly more accurate forecasts than AR(1) dynamics.

As pointed out, in this study we are particularly interested in the forecasting

		3m	6m	12m	2y	3y	5y	7y	10y
	RW	21.9	21.5	22.1	24.0	26.1	27.7	30.0	27.9
	NSAR	1.47*	**0.97**	**0.92**	1.04	1.14*	1.21*	1.07	1.03
	NSVAR	1.12	**0.90**	1.07	1.08*	1.06	1.13*	1.04	1.04
h=1	PCAAR	1.16*	**0.95**	1.01	1.13*	1.06*	1.02	1.01	1.00
	PCAVAR	1.00	**0.88**	1.02	1.09*	1.03	1.03	1.04	**1.00**
	AR1	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
	RW	81.1	82.6	79.8	72.7	72.4	70.4	72.5	66.9
	NSAR	1.23	1.11	1.08	1.20	1.26*	1.28*	1.16	1.06
	NSVAR	**0.99**	**0.97**	**0.98**	1.05	1.06	1.09	1.02	**0.97**
h=6	PCAAR	1.03	1.03	1.04	1.09	1.07	1.03	**0.98**	**0.94”**
	PCAVAR	**0.98**	1.01	1.03	1.07	1.05	1.04	1.02	**0.96**
	AR1	1.04	1.03	1.03	1.01	1.01	1.04	1.05	1.05
	RW	145.0	144.0	133.4	112.5	100.7	86.7	85.7	78.1
	NSAR	1.17	1.12	1.14	1.30	1.43	1.55*	1.42	1.31
	NSVAR	**0.95**	**0.95**	**0.98**	1.08	1.14	1.20	1.12	1.04
h=12	PCAAR	**0.95**	**0.97**	**1.00**	1.06	1.09	1.09	1.02	**0.94**
	PCAVAR	**0.96**	**0.99**	1.03	1.10*	1.13	1.15	1.11	1.03
	AR1	1.05	1.04	1.03	1.00	1.02	1.12	1.18	1.20

Table 2. Forecasting results of US yields for h=1, h=6 and h=12 months-ahead forecasting horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. We report root mean squared error (RMSE) for the out-of-sample period **2004:1 - 2013:12** (N = 96). The first line reports the RMSE for the random walk (expressed in basis points). The RMSEs of all other models are expressed relative to the random walk. Hence, numbers smaller than one (**reported in bold**) indicate that models outperform the random walk. Numbers larger than one indicate inferior performance. Numbers larger than one indicate inferior performance. (”) indicates statistical significant forecasting superiority of the respective models against the random walk measured by the DM-statistic on a 5% or smaller significance level. (*) indicates statistical significant forecasting inferiority against the random walk. The DM-statistics are reported in Appendix A. See section 4 for a description of the selected models.

performance of the models for the low interest rate environment following the GFC. Unfortunately, the RMSE does not provide any insights for which

particular time periods the models perform well and poor, since it only measures the global forecasting performance over the entire out-of-sample period. Thus, information about the dynamic forecasting performance throughout the forecasting period is lost.
To reveal the dynamic forecasting performance we take a closer look at the development of the forecasting accuracy through time. First, we construct sequences of local relative RMSEs based on rolling windows throughout the forecasting period. Second, we divide the forecasting period into sub-samples to conduct a sub-sample analysis.

5.3 Dynamic forecasting evaluation

Based on the forecasting errors computed in the forecasting exercise above, we define a dynamic relative RSME as the sequence of local relative RMSEs over centered rolling windows of size p (assuming p to be an even number) for $t* = R + p/2 ... T - p/2 + 1$. The intention of this innovative measure is to look at the entire time path of the models relative forecasting performance. For each model and the random walk the local RMSE for the respective rolling window is given by

$$RMSE^{m,local}_{t*,\tau,h} = \sqrt{\frac{1}{p} \sum_{j=t-p/2}^{t=p/2-1} (\hat{y}^m_{t+h/h,\tau} - \hat{y}^m_{t+h,\tau})^2}. \tag{14}$$

We then express the sequence of local $RMSEs^{m,local}_{t*,\tau,h}$ for all models relative to the random walks local $RMSEs^{RW,local}_{t*,\tau,h}$ sequence. As indicated above, values smaller than one indicate that models outperform the random walk. Values larger then one indicate inferior forecasting performance against the random walk. In Figure 2 we plot the local relative RMSEs for a twelve months forecast horizon and selected short, medium and long maturities.

The dynamic forecast evaluation reveals that prior and during the GFC all models compete relatively closely with the random walk for all maturities with some periods of outperformance and some periods of underperfor-

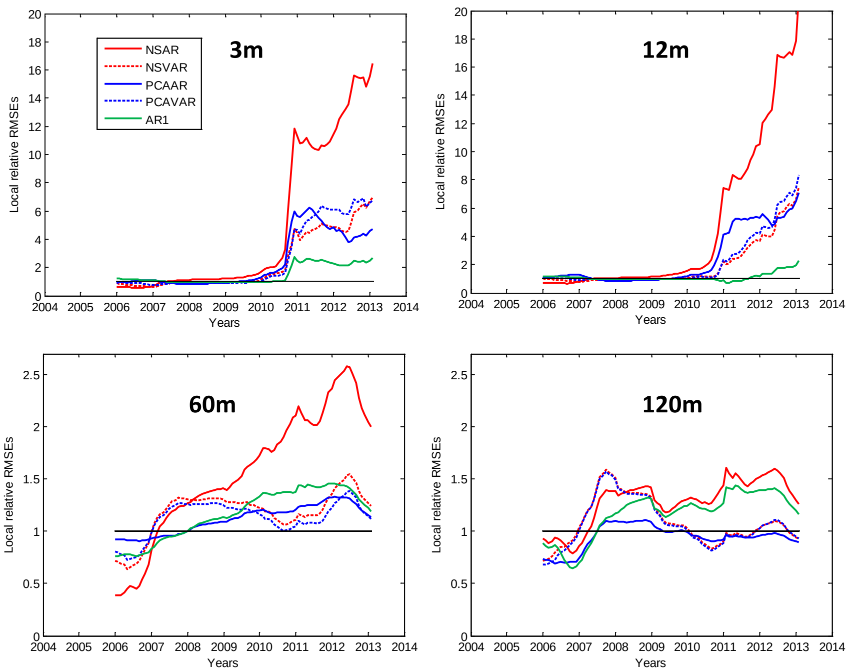


Figure 2. Dynamic relative three-months, twelve-months, 60-months and 120-months yield RMSEs for all models against the random walk for a h=12 forecast horizon. For each model and the random walk we calculate sequences $(t* = R + p/2...T - p/2 + 1)$ of local RMSEs for rolling windows of size p=24 throughout the the forecasting period from 2004:01 - 2013:12. We then calculate the dynamic relative RMSE by expressing the sequence of local $RMSEs^{m}_{t*,\tau,h}$ for each model relative to the random walks local $RMSE^{RW}_{t*,\tau,h}$ sequence. Hence, values smaller than one indicate that models outperform the random walk. Values larger then one indicate inferior forecasting performance against the random walk.

mance. For the ten year yield, this close race also lasts throughout the entire forecasting period. For the short an medium yields however, things change dramatically subsequent to the GFC. The forecasting accuracy worsens significantly in relative terms for all factor models. The Nelson-Siegel model with AR(1) factor dynamics fares particularly bad. While the AR(1) process performs better than the factor models it also is consistently dominated by the random walk from 2010 onwards.

5.4 Subsample Analysis

These conclusions are confirmed by the results of the sub-sample analysis reported in Table 3. For the first sub-sample from 2004:01-2009:12 the results are roughly in line with the results reported for the entire sample above. The factor models are partly able to beat the random walk especially for short and medium yields. However, the outperformance is not statistically significant. Absolute RMSEs of the sub-sample are generally high as all models and the random walk struggle to predict the sudden drop in yields during the GFC. The results for the crucial sub-sample period after the GFC (2009:01-2013:10) are striking. In absolute terms, the RMSEs drop notably. In relative terms, the forecasting accuracy of the selected dynamic factor models for short and medium term yields worsens significantly compared to the random walk. For some of the models, calculated RMSEs are even more than ten times higher than the random walk. The poor forecasting performance of the considered models relative to the random walk is particularly pronounced for shorter and medium maturities, i.e. three-months, six-months and 12-months yields.

Moreover, conducted Diebold-Mariano tests[9] show that the considered models are significantly outperformance by the RW for many maturities and forecasting horizons, often even at the 1% level. Unreported analysis confirms, that similar results hold against the AR(1) model. In other words, after the GFC the random walk and a simple first order autoregressive process are able to significantly outperform all considered dynamic factor model variations. Naturally, it is expected that the forecasting performance of forecasting models varies over time. However, this dimension of outperformance is still a striking result for such an extended period.

There are different reasons for the poor performance of the applied econometric models against a simple random walk during the post GFC low yield and relative flat interest rate environment. One reason may be that the models are calbrated over a time period that also includes a dynamic behavior of the term structure of the yield curve as well significant changes in interest rates for given maturities. The estimated models may then overstate the dynamics

[9]Detailed results and test statistics for the conducted tests are reported in Appendix A.

		3m	6m	12m	2y	3y	5y	7y	10y
		2004:01 - 2008:12							
	RW	30.8	30.1	30.5	31.4	32.3	30.8	32.7	27.9
h=1	NSAR	1.26*	**0.92”**	**0.91**	1.04	1.08	1.11	1.07	1.01
	NSVAR	1.06	**0.83**	**0.93**	1.04	1.05	1.09*	1.05	1.00
	PCAAR	1.14	**0.95**	**0.99**	1.09*	1.05	1.02	1.04	1.01
	PCAVAR	**0.98**	**0.82”**	**0.96**	1.07	1.03	1.02	1.06	1.02
	AR1	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02
	RW	113.0	114.0	108.5	97.8	92.3	72.9	65.9	53.7
h=6	NSAR	1.02	**0.95**	**0.93**	**1.00**	1.01	1.09	1.10	1.08
	NSVAR	**0.92**	**0.92**	**0.96**	1.04	1.05	1.11	1.10	1.07
	PCAAR	**0.97**	**0.98**	1.00	1.03	1.00	1.00	1.00	**0.97**
	PCAVAR	**0.92**	**0.97**	1.02	1.07	1.05	1.07	1.10	1.06
	AR1	1.04	1.04	1.03	1.00	**0.99**	**0.98**	1.02	1.04
	RW	207.0	204.3	186.6	156.8	136.5	100.0	83.9	63.6
h=12	NSAR	**0.97**	**0.93**	**0.92**	**0.99**	1.01	1.10	1.15	1.21
	NSVAR	**0.90**	**0.90**	**0.95**	1.05	1.08	1.16	1.17	1.21
	PCAAR	**0.87**	**0.90**	**0.93**	**0.98**	**0.98**	**0.99**	1.00	**0.98**
	PCAVAR	**0.92**	**0.96**	1.01	1.08	1.10	1.14	1.18	1.19
	AR1	1.05	1.05	1.03	**0.98**	**0.96**	**0.96**	1.02	1.08
		2009:01 - 2013:12							
	RW	4.1	4.7	7.4	12.9	18.0	24.6	28.1	28.7
h=1	NSAR	5.83*	2.22*	1.16	1.06	1.31*	1.35*	1.09	1.04
	NSVAR	2.95*	2.34*	2.45*	1.30*	1.06*	1.19*	1.03	1.07
	PCAAR	1.93*	1.25	1.28*	1.30*	1.10*	1.03	**0.99**	**0.99**
	PCAVAR	1.82*	2.22*	1.69*	1.21*	1.03	1.05	1.01	**0.98**
	AR1	1.11	1.08	1.02	1.01	1.01	1.01	1.00	1.00
	RW	7.9	9.4	9.9	22.4	38.9	63.7	75.9	76.2
h=6	NSAR	9.19*	6.74*	5.87*	2.91*	2.10*	1.47	1.18	1.03
	NSVAR	4.35*	3.40*	2.67*	**0.92**	1.04	1.04	**0.95**	**0.89**
	PCAAR	3.57*	3.07*	3.19*	1.84*	1.34*	1.04	**0.95**	**0.91”**
	PCAVAR	4.32*	3.59*	2.87*	1.05	**0.98**	**0.96**	**0.94**	**0.89**
	AR1	1.59	1.33	1.04	1.07	1.09	1.07	1.06	1.04
	RW	8.9	10.5	11.9	26.2	43.8	74.5	92.7	93.6
h=12	NSAR	13.15*	10.65	9.41	4.76	3.24*	2.03*	1.55*	1.31
	NSVAR	5.43	4.27	3.14*	1.37	1.36*	1.18	1.02	**0.92**
	PCAAR	5.28*	4.71*	4.47*	2.33*	1.65*	1.17	**1.00**	**0.91**
	PCAVAR	5.46*	4.52*	3.43*	1.42	1.24	1.08	**1.00**	**0.91**
	AR1	2.50*	2.00	1.10	1.19	1.28*	1.27*	1.24	1.20

Table 3. Sub-sample forecasting results of US yields for h=1, h=6 and h=12 months-ahead forecasting horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. We report root mean squared error (RMSE) for the sub-sample periods from **2004:01-2008:12** and **2009:01-2013:12**. The first line reports the RMSE for the random walk (expressed in basis points). The RMSEs of all other models are expressed relative to the random walk. Hence, numbers smaller than one (**reported in bold**) indicate that models outperform the random walk. Numbers larger than one indicate inferior performance. Numbers larger than one indicate inferior performance. (”) indicates statistical significant forecasting superiority of the respective models against the random walk measured by the DM-statistic on a 5% or smaller significance level. (*) indicates statistical significant forecasting inferiority against the random walk. The DM-statistics are reported in Appendix A. See section 4 for a description of the selected models.

of individual yiels as well as for the entire yield curve during the unique low interest rate period from 2009 to 2013. Further, since the models are estimated during periods when short-term interest rates were significantly higher than after the GFC, created forecasts may not only overstate the dynamics of the interest rate term structure but possibly also the levels of short-term interest rates.

5.5 Sensivity of results towards forcast evaluation metrics

These results obviously raise the question, why the poor relative forecasting performance for short and medium yields subsequent to the GFC is not fully reflected in the results reported for the entire forecasting period. After all, the critical time period makes up half of the out-of-sample period. This is also highly important for future yield curve forecasting studies encompassing this time period.
The answer can be found in the decreasing magnitude of forecasting errors caused by the low interest rate environment after the GFC. Not surprisingly, with flat short and medium yields close to the zero bound, forecasting errors and RMSEs drop significantly in absolute terms. This is illustrated in Figure 3, where we plot the six-months yield forecasts against the six-months actual yield for the forecasting horizons h=12 and the corresponding forecasting errors for the random walk, one Nelson-Siegel and one PCA variation.

First of all, it is quite obvious that, different to random walk and AR(1) model, all selected dynamic factor models have problems forecasting the period after January 2010. While AR(1) model and random walk adapt rather quickly to the changed environment both factor models, especially the parametric Nelson-Siegel model with AR(1) factor dynamics, continuously over- and under predict the actual yield. Only the PCAAR model picks up the new interest rate environment towards the end of the period. It is also important to note, that at times all models predict negative yields when the actual yield is close to the zero bound. This is a highly undesired effect for many pricing and hedging purposes.

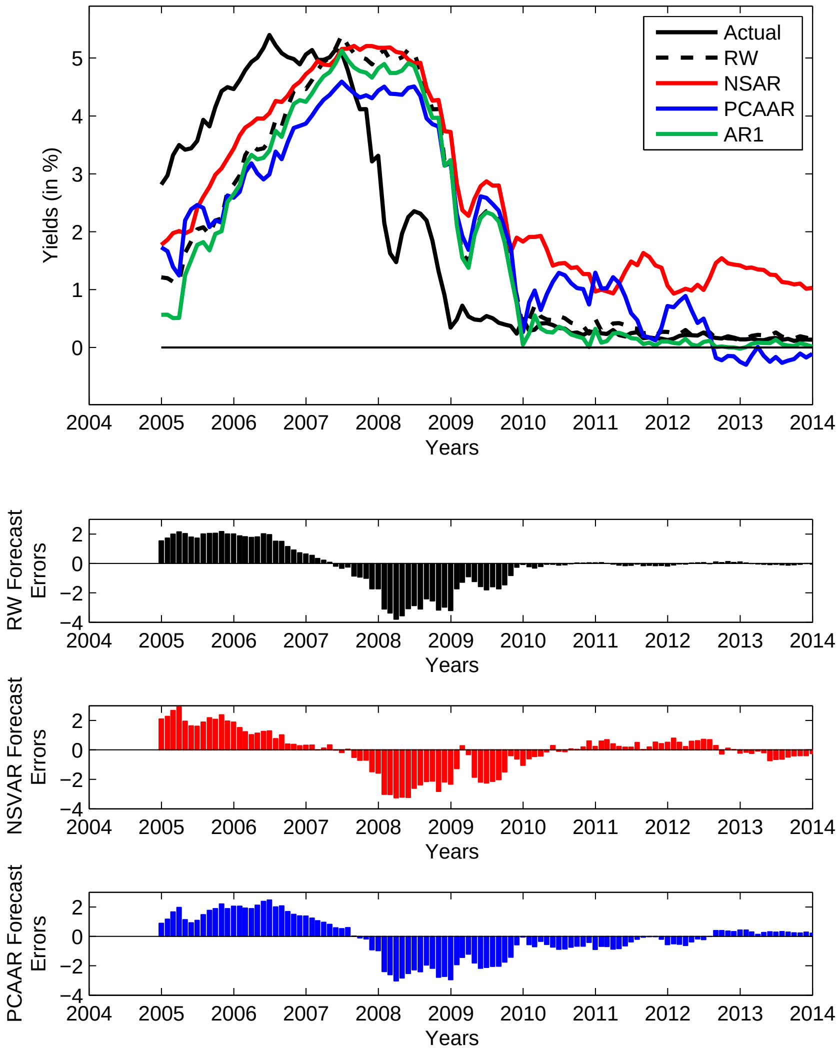


Figure 3. US yield forecasts and forecasting errors for randwom walk, NSAR, PCAAR and AR1 model. We plot the six-months actual yield together with the forecasts of the selected models for a forecasting horizon of h=12 months (upper panel). The lower panels provide plots of the time series of the forecasting errors (Actual yield - forecasted yield) for the random walk, NSAR and PCAAR model.

Figure 3 also confirms that the forecasting errors for small and medium yields become rather small in absolute terms subsequent to the GFC. Usually the forecast errors especially for shorter maturities are relatively large as the shorter maturities are rather volatile. Thus the poor relative forecasting performance after the GFC vanishes in global forecast evaluation measures averaged over the entire forecasting period. The RMSE being based on a quadratic loss function further aggravates this effect. What usually is a desired outcome may lead in this case to biased inferences about the forecasting performance post the GFC. This highlights one of the most important points of this paper: investigating the global (or average) absolute forecasting performance may hide important information about the relative forecasting performance over time and lead to false inferences of the true forecasting capabilities of models.

Interestingly, the unique behaviour of yields after the GFC also poses a challenge for other forecasting measures relying on absolute differences in forecasting errors. As an additional evaluation metric we apply the innovative fluctuation test developed by Giacomini and Rossi (2010). This test allows to look at the entire path of local test-statistics and reveals the statistical significance of the forecasting performance over time. The sequence of local test-statistics is calculated based on the local loss function differentials ΔL_j computed over centered rolling windows of size p and given as

$$F^m_{t*,\tau,h} = \hat{\sigma}^{-1} p^{-1/2} \sum_{j=t-p/2}^{t+p/2-1} \Delta L_j(\hat{y}^{RW}_{t+h/h,\tau}, \hat{y}^m_{t+h/h,\tau}), \tag{15}$$

where $\hat{\sigma}^2$ is a HAC estimator of the asymptotic (long-run) variance. The test statistic $F^m_{t*,\tau,h}$ is equivalent to Diebold and Marianos (1995) computed over rolling windows. Giacomini and Rossi (2010) also provide critical values to test the null of equal predictive accuracy. See Giacomini and Rossi (2010) for more details.

In Figure 4 we plot the fluctuation test statistics for the six-months yield and a forecast horizon of h=12 based on rolling windows of size p=24 with corresponding two-sided critical values.[10] The fluctuation test correctly reflects the direction of out- and underperformances. However, none of the local test-

[10]Note that unlike the forecasting exercises conducted in previous sections that were based on a recursive window estimation, flucuations tests were conducted based on a rolling window estimation as proposed in Giacomini and Rossi (2010). We also conducted the fluctuation tests using a recursive window estimation, where results did not differ qualitatively from the rolling window methodology.

statistics post the GFC indicates statistically significant outperformance by the random walk. This is surprising, given the statistically significant outperformance for the sub-sample period reported in Table 2. Further analysis reveals, that the decreasing loss functions distort the local test-statistics calculated based on the the global $\widehat{LRV}_{\bar{d}}$. This confirms the observation of Martins and Perron (2012) who find power problems of the fluctuation tests in the presence of instabilities in the differences of the loss functions.

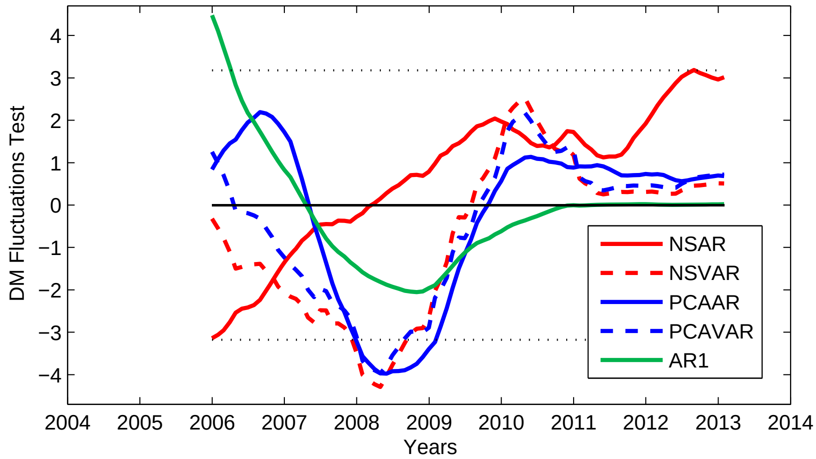


Figure 4. Fuctuations test statistics for all models against the random walk for six-months yields and h=12 forecast horizon. The $t* = R + p/2 ... T - p/2 + 1$ sequence of local test-statistics is calculated based on rolling windows of size p=24 throughout the the forecasting period from **2004:01 - 2013:12**. Values smaller than zero zero indicate that models outperform the random walk. Values larger then zero indicate inferior forecasting performance against the random walk. Values larger/smaller than the critical values indicated statistically significance. The critical values [3.01; −3.01] are obtained from Giacomini and Rossi (2010).

6 Forecast Combination

A natural question to ask is how to best approach the instability in the relative performance of the selected models. Previous research discusses several interesting measures to approach unstable environments, for example adaptive forecasting (Chen and Niu, 2014) or regime switching models (Xiang and Zhu, 2013). A promising approach advocated in recent literature is also to combine the forecasts of individual models. Several studies (Guidolin and Timmermann, 2009; Pooter et al., 2010) have shown that combining multiple forecasts may increase the forecasting accuracy. This approach is particularly promising in our case, since the forecasting accuracy of our selected models heavily varies over time, often diametrically. Thus combined forecasts are likely to be more robust to structural instability than either of the individual models.

In the following section we therefore investigate different combinations of individual forecasts in order to improve the forecasting accuracy especially for the crucial time period after the GFC. We consider three different forecast combination strategies. The first simply includes all four factor models (NSAR, NSVAR, PCAAR, PCAVAR - 'factor'). The second one includes the NSAR, PCAAR and the AR1 forecast ('fAR1'). The graphical analysis in Figure 3 shows, that NSAR and PCAAR model seem to be diametrically biased in their forecasts of short and medium yields after the GFC. Combining both models should thus improve the individual forecasts. Including the AR1 forecast is an obvious choice as the AR(1) process performs rather well compared to the random walk. For the third one we also include forecasts generated by a random walk no-change model (NSAR, PCAAR, AR1, RW - 'far1RW') into the combination scheme. Given the superior forecasting performance of the random walk in particular after the GFC, it is reasonable to expect that combining forecasts that also include a simple random walk model will further improve the results for the second sub-sample period from 2009:01-2013:10.

With M models and hence M individual forecasts for a τ-maturity yield at time t a linearly combined forecast ("cf") based on weights $w^{\tau}_{t,m}$ is given by

$$\hat{y}^{cf}_{t+h/t,\tau} = w^{\tau\prime}_{t}\hat{y}^{m}_{t+h/t,\tau} = \sum_{m=1}^{M} w^{\tau}_{t,m}\hat{y}^{m}_{t+h/t,\tau}, \tag{16}$$

where the $Mx1$ vector of weights w^{τ}_{m} is time varying.

For each forecast combination we consider two forecasting combination schemes: equal weights (**CFEW**) and performance weights (**CFPW**). For equal weights, each weight is given by

$$w^{\tau}_{t,m} = 1/M. \tag{17}$$

For performance weights, each forecast is weighed by the inverse of its MSE (Mean squared error)[11] over the previous $v = 24$ months. The MSE for each model m, maturity τ at time t is calculated as

$$MSE^{\tau}_{t,m} = \frac{1}{v} \sum_{t-v}^{t} e^2_{t+h/t,m}, \tag{18}$$

where $e^2_{t+h/t,m}$ is the squared forecast error of model m at time t. Each weight is then given as

$$w^{\tau}_{t,m} = \frac{1/MSE^{\tau}_{t,m}}{\sum_{m=1}^{M} 1/MSE^{\tau}_{t,m}}. \tag{19}$$

This way, a model with a previously lower MSE is given a relatively larger weight than a model with a previously higher MSE performing model. Combining the three forecast combination strategies with the two combination schemes leaves us with six forecast combination strategies which we denote **CFEWfactor**, **CFPWfactor**, **CFEWfar1**, **CFPWfar1**, **CFEWfar1RW** and **CFPWfar1RW**.

Table 4 presents the results of these six forecasting strategies for the entire sample period. The corresponding Diebold-Mariano statistics are reported in Appendix B. As indicated by the results, still none of the combination strategies is able to consistently beat the random walk on a statistically significant level. However, combining forecasts does indeed reduce the model uncertainty and delivers more stable forecasts across maturities and forecasting horizons. More importantly, most forecasting strategies perform better than the individual dynamic factor models in Table 2. Including the random walk into the combination strategy does not significantly improve the performance. We also find that the performance based combination schemes deliver slightly more accurate forecasts than the equally weighted performance schemes.

Next, we focus on the forecasting performance within the sub-sample periods, see Table 5. For the period from 2004:01 to 2008:12 combining individual forecasts also slightly improves the forecasting accuracy compared

[11]Note that we follow Timmermann (2006) in using the MSE to construct the weights instead of the RMSE we report in Tables 2 and 3.

		3m	6m	12m	2y	3y	5y	7y	10y
	RW	21.9	21.5	22.1	24.0	26.1	27.7	30.0	27.9
h=1	CFEWfactor	1.08	**0.86”**	**0.93**	1.05*	1.04	1.05*	1.05	1.01
	CFPWfactor	1.05	**0.85”**	**0.92**	1.04*	1.04	1.04*	1.05	1.00
	CFEWfar1	1.11*	**0.95**	**0.96**	1.04*	1.04	1.05	1.02	**1.00**
	CFPWfar1	1.05	**0.94**	**0.95**	1.03*	1.03	1.03	1.01	**0.99**
	CFEWfar1RW	1.07	**0.96**	**0.96**	1.03*	1.03	1.03	1.01	**1.00**
	CFPWfar1RW	1.02	**0.95**	**0.95**	1.02*	1.02	1.02	1.00	**0.99**
	RW	81.1	82.6	79.8	72.7	72.4	70.4	72.5	66.9
h=6	CFEWfactor	**0.92**	**0.92**	**0.95**	1.01	1.01	1.05	1.06	1.03
	CFPWfactor	**0.88**	**0.89**	**0.91**	**0.98**	**0.98**	1.03	1.05	1.02
	CFEWfar1	1.02	**1.00**	**1.00**	1.05	1.06	1.08	1.04	**1.00**
	CFPWfar1	**0.95**	**0.95**	**0.95**	1.00	1.02	1.06	1.03	**0.99**
	CFEWfar1RW	1.01	**1.00**	**0.99**	1.03	1.04	1.04	1.02	**0.99**
	CFPWfar1RW	**0.95**	**0.96**	**0.96**	**0.99**	**1.00**	1.02	1.01	**0.98**
	RW	145.0	144.0	133.4	112.5	100.7	86.7	85.7	78.1
h=12	CFEWfactor	**0.89**	**0.90**	**0.93**	1.00	1.02	1.08	1.11	1.13
	CFPWfactor	**0.86**	**0.87**	**0.90**	**0.96**	**0.97**	1.04	1.09	1.10
	CFEWfar1	**0.99**	**0.98**	**1.00**	1.07	1.13	1.21	1.18	1.13
	CFPWfar1	**0.93**	**0.94**	**0.95**	**1.00**	1.04	1.14	1.13	1.07
	CFEWfar1RW	**0.99**	**0.98**	**0.99**	1.04	1.08	1.14	1.11	1.07
	CFPWfar1RW	**0.94**	**0.95**	**0.95**	**0.98**	**1.00**	1.06	1.06	1.03

Table 4. Forecasting combination results of US yields for h=1, h=6 and h=12 months-ahead forecasting horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. We report the root mean squared error (RMSE) for the out-of-sample period **2004:1 - 2013:12** (N = 96). The first line reports the RMSE for the random walk (expressed in basis points). The RMSEs of all other models are expressed relative to the random walk. Hence, numbers smaller than one (**reported in bold**) indicate that models outperform the random walk. Numbers larger than one indicate inferior performance. (”) indicates statistical significant forecasting superiority of the respective models against the random walk measured by the DM-statistic on a 5% or smaller significance level. (*) indicates statistical significant forecasting inferiority against the random walk. The DM-statistics are reported in Appendix B. See text for a description of the selected forecast combination strategies.

to the previous results. For several short term yields and in particular for a $h = 12$ months forecasting horizon the outperformance is even statistically significant. All three strategies fare comparably well. Again, there is no notable advantage by including the random walk. Interestingly, there is also no notable difference between equally weighted and performance weighted combination schemes.

For the crucial second sub-sample period after the GFC (2009:01 to 2013:12), combining different models significantly improves the forecasting performance, albeit most of the strategies are still being dominated by the random walk. The RMSE for a three-months yield forecast over a six-months horizon, for example, decreases to 1.88 relative to the forecasting error of a random walk for the performance weighed combination of all factor models (CFPWfactor). Recall that the initial RMSEs for the individual models in Table 3 range from 3.57 to 9.19 for the same maturity and forecasting horizon. In particular the CFPWfar1 strategy performs comparably well with the relative RMSEs being significantly smaller than the individual RMSEs for this period. Obviously this is partly due to the relatively good performance of the simple AR(1). Not surprisingly, the most promising strategy turns out to be the performance weighted forecast combination of the NSAR and PCAAR model with both the AR(1) model and the random walk (CFPWfarRW). This strategy even outperforms the simple random walk forecast for most forecast horizons and maturities, in particular for the 3m, 6m and 12m yields as well as for yields with longer maturities such as 7y and 10y yields.

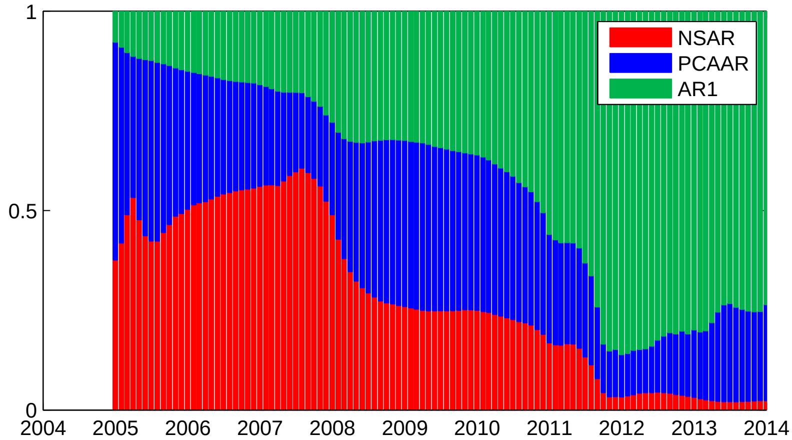
In general, weighting the individual models based on their previous performance makes a remarkable difference compared to the equally weighted forecast combination for this time period. Further examining this issue, we investigate the weights allocated to each of the included models, when the performance based weighting technique is applied to create forecast combinations. Figure 5 displays the development of the weights for the two most promising performance weighted forecast combination strategies, CFPWfar1 and CFPWfarRW. The figure illustrates that the AR(1) process (forCFPWfar1) and the random walk (for CFPWfarRW) receive consistently high

		3m	6m	12m	2y	3y	5y	7y	10y
		2004:01 - 2008:12							
	RW	30.8	30.1	30.5	31.4	32.3	30.8	32.7	27.9
h=1	CFEWfactor	1.08	**0.86”**	**0.93**	1.05	1.04	1.05	1.05	1.01
	CFPWfactor	1.05	**0.85”**	**0.92**	1.04	1.04	1.04	1.05	1.00
	CFEWfar1	1.08	**0.95**	**0.95**	1.03*	1.03	1.03	1.04	1.01
	CFPWfar1	1.05	**0.94”**	**0.94**	1.02*	1.02	1.03	1.03	1.00
	CFEWfar1RW	1.05	**0.96**	**0.96**	1.02*	1.02	1.02	1.02	1.00
	CFPWfar1RW	1.01	**0.95**	**0.95**	1.02*	1.01	1.02	1.02	**1.00**
	RW	113.0	114.0	108.5	97.8	92.3	72.9	65.9	53.7
h=6	CFEWfactor	**0.92**	**0.92**	**0.95**	1.01	1.01	1.05	1.06	1.03
	CFPWfactor	**0.88**	**0.89**	**0.91”**	**0.98**	**0.98**	1.03	1.05	1.02
	CFEWfar1	**0.98**	**0.97**	**0.97**	**0.99**	**0.98**	1.01	1.03	1.01
	CFPWfar1	**0.92**	**0.93**	**0.93”**	**0.95**	**0.95**	**0.99**	1.02	1.00
	CFEWfar1RW	**0.98**	**0.98**	**0.98**	**0.99**	**0.98**	**1.00**	1.01	**1.00**
	CFPWfar1RW	**0.93**	**0.94**	**0.94**	**0.95**	**0.95**	**0.98**	1.01	**0.99**
	RW	207.0	204.3	186.6	156.8	136.5	100.0	83.9	63.6
h=12	CFEWfactor	**0.89”**	**0.90”**	**0.93”**	1.00	1.02	1.08	1.11	1.13
	CFPWfactor	**0.86”**	**0.87”**	**0.90”**	**0.96**	**0.97**	1.04	1.09	1.10
	CFEWfar1	**0.94”**	**0.94**	**0.95**	**0.96**	**0.97**	**1.00**	1.04	1.07
	CFPWfar1	**0.89”**	**0.90”**	**0.90”**	**0.92**	**0.92**	**0.96**	1.03	1.04
	CFEWfar1RW	**0.95**	**0.95**	**0.96**	**0.97**	**0.97**	**0.99**	1.02	1.03
	CFPWfar1RW	**0.91**	**0.92**	**0.92**	**0.93**	**0.92**	**0.94**	1.00	1.02
		2009:01 - 2013:12							
	RW	4.1	4.7	7.4	12.9	18.0	24.6	28.1	28.7
h=1	CFEWfactor	2.19*	1.33*	1.43*	1.14*	1.07	1.12	1.02	1.01
	CFPWfactor	1.27	1.05	1.22	1.10	1.05	1.09	1.01	1.00
	CFEWfar1	2.26*	1.20*	1.07	1.06	1.07	1.07	1.00	**0.99**
	CFPWfar1	1.01	**0.98**	1.00	1.03	1.05	1.04	**0.99**	**0.99**
	CFEWfar1RW	1.82*	1.11	1.03	1.04	1.05	1.04	**0.99**	**0.99**
	CFPWfar1RW	**0.97**	**0.94**	**0.96**	1.01	1.03	1.02	**0.97**	**0.99**
	RW	7.9	9.4	9.9	22.4	38.9	63.7	75.9	76.2
h=6	CFEWfactor	2.78*	2.18*	1.99*	1.30*	1.24*	1.08	**0.98**	**0.91**
	CFPWfactor	1.88*	1.69*	1.69*	**0.94**	1.05	1.02	**0.95**	**0.90**
	CFEWfar1	3.13*	2.49*	2.43*	1.62*	1.35*	1.12	1.03	**0.97**
	CFPWfar1	1.19	1.09	1.07	1.27*	1.22	1.09	1.00	**0.96**
	CFEWfar1RW	2.49*	2.04*	2.02*	1.43*	1.23*	1.07	1.00	**0.97**
	CFPWfar1RW	**0.98**	**0.95**	**1.00**	1.14	1.11	1.03	**0.97**	**0.96**
	RW	8.9	10.5	11.9	26.2	43.8	74.5	92.7	93.6
h=12	CFEWfactor	4.12*	3.50*	3.34*	2.09*	1.74*	1.32*	1.11	**0.99**
	CFPWfactor	2.73*	2.47*	2.35*	1.41*	1.37	1.17	1.03	**0.94”**
	CFEWfar1	4.34*	3.73*	3.72*	2.44*	1.90*	1.42*	1.22	1.11
	CFPWfar1	1.42	1.29	1.01	1.55*	1.55*	1.30	1.15	1.04
	CFEWfar1RW	3.39*	2.98*	3.00*	2.05*	1.64*	1.28	1.14	1.06
	CFPWfar1RW	**0.93**	**0.80**	**0.82**	1.27*	1.27*	1.15	1.06	1.00

Table 5. Sub-sample forecasting combination results of US yields for h=1, h=6 and h=12 months-ahead forecasting horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. We report the root mean squared error (RMSE) for the out-of-sample periods **2004:1 - 2008:12** and **2009:1 - 2013:12** . The first line reports the RMSE for the random walk (expressed in basis points). The RMSEs of all other models are expressed relative to the random walk. Hence, numbers smaller than one (**reported in bold**) indicate that models outperform the random walk. Numbers larger than one indicate inferior performance. (”) indicates statistical significant forecasting superiority of the respective models against the random walk measured by the DM-statistic on a 5% or smaller significance level. (*) indicates statistical significant forecasting inferiority against the random walk. The DM-statistics are reported in Appendix B. See text for a description of the selected forecast combination strategies.

weights when producing the combined forecasts. The figure also illustrates how in more recent periods, the weight of the AR(1) and random walk significantly increase due to a superior forecasting performance. As illustrated in the lower panel of Figure 5, for the initial forecasting periods from 2005-2007, forecasts created by the PCA and Nelson-Siegel based factor models obtain relatively high weights, while from 2010 onwards the random walk becomes the dominant model and crowds out the factor models but also the AR(1) process.

Overall, our results clearly illustrate that forecast combinations are able to provide superior forecasts for the term stucture of interest rates in comparison to using individual econometric models. We also find strong evidence for the fact that during separate regimes of yield curve behavior, different models will provide the most appropriate forecasts. In particular during the transition from a more volatile behavior of the yield curve to the current low interest rate environment with only minor fluctuations, weights allocated to the individual models change dramatically. Therefore, our results strongly encourage the use of forecast combination schemes, with a random walk no-change model as one of the included models.



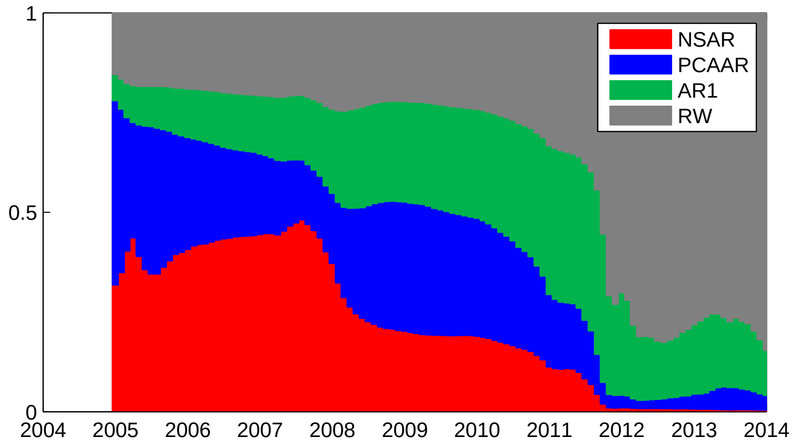


Figure 5. Development of forecast combination weights. We plot the changes in the weights of the CFPWfar1 (top plot) and CFPWfar1RW (bottom plot) strategy for six-monhts maturity and h=12 forecast horizon. The CFPWfar1 strategy encompasses the NSAR, PCAAR and AR1 model. The CFPWfar1RW includes in addition the random walk. The Weights are calculated based on the inverse MSE of the previous v=24 months. See text for a more detailed description of the selected forecast combination strategies.

7 Conclusion

This paper provides a pioneer study in documenting the challenge which the current low interest rate environment poses to popular dynamic factor yield curve forecasting models. To examine the forecasting accuracy during this unique time period we apply a dataset of monthly US Treasury yields (12 maturities ranging from three-months to 120-months) obtained from Bloomberg

for the time period from 1995:01 to 2013:12. We focus on the popular class of dynamic factor models and investigate variations of the parametric dynamic Nelson-Siegel model and regressions on principal components (PCA).
The forecasting results for the entire period confirm findings from previous forecasting studies. RMSEs are generally smaller for longer term maturities and the forecasting performance worsens with longer forecasting horizons. The selected factor models perform relatively well for short term maturities, but all models fail to consistently beat the random walk.
In our study, we are particular interested in the forecasting performance of the estimated models subsequent to the GFC period (2009:01-2013:12) that is dominated by flat, non-volatile short and medium interest rates. Given this unique interest rate environment, we would expect the random walk to perform relatively well in comparison to the applied econometric models during this sub-period. However, we argue that this behavior will not be detected by examining forecasting errors over the entire sample period, since such an analysis does not reveal when individual models make their largest and smallest forecast errors. We therefore conduct a dynamic forecasting evaluation and sub-sample analysis. As it turns out the relative forecasting accuracy for short- and medium term rates changes dramatically after the GFC. The investigated dynamic factor models not only fail to beat the random walk but are completely outperformed in relative terms. Diebold-Mariano statistics show that the outperformance of the models by a simple random walk no-change forecast is also statistically significant, often at the 1% level.
This naturally raises the question, why these results were not reflected in the RMSEs reported for the entire period. The answer lies in the size of the forecasting errors after the crisis. As the forecast errors become relatively small in absolute terms after the GFC, the results of this period that represents half of our out-of-sample forecasting period, contribute relatively little to the total, and thus, also to the average forecasting error, that is measured by the RMSE. Investigating only the global forecasting performance, therefore, may hide important information about the relative forecasting performance of competing models over time.

Overall, the above results for the period after the GFC are startling. It is well known that model uncertainty in regard to methodology, time period and dataset is generally high when forecasting the term structure of interest rates (Moench, 2008; Pooter et al., 2010). Naturally, the performance of forecasting models varies over time, albeit the magnitude of the relative outperformance over such a prolonged period is striking. We argue that since the applied dynamic models are typically calibrated over a sample period that also includes significant changes in interest rate as well as volatile periods for the term structure of the yield curve, they may overstate the dynamics of individual yiels as well as for the entire yield curve during the unique low interest rate period from 2009 to 2013. Moreover, the models were also estimated during periods when interest rates were significantly higher than during the post GFC period such that forecasts created by the applied models will not only overstate the dynamics of the interest rate term structure but possibly also interest rate levels what is also evidenced by our results. As this unique interest rate environment may well last for some more time into the future[12] the above results have important implications for current and future yield curve forecasting exercises.

First, it is crucial to carefully examine the dynamic behaviour of the term structure and conduct sub-sample analysis accordingly. It is still common to measure forecasting accuracy predominantly with RMSEs computed over the entire sample period and select the model with the best global forecasting performance. However, a thorough sub-sample analysis and dynamic forecasting measures are crucial to truly expose a model's predictive abilities. Dynamic forecast evaluation measures such as, e.g., fluctuation tests suggested by Giacomini and Rossi (2010) are required to identify periods of superior or inferior forecasting accuracy and should. However, as illustrated in our study, even such tests, focusing on the local performance of forecasting models, may have difficulties in significantly detecting differences in the perofrmance between the applied techniques in the presence of instabilities

[12] Fed chair Janet Yellen only recently confirmed there will be 'considerable time' before the central bank may raise its benchmark rate. See the transcript of Chair Yellen's Press Conference on 19 March, 2014.

in the differences of the loss functions (Martins and Perron, 2012).
Secondly, future yield curve forecasting studies need to pay special attention to the forecasting accuracy of the investigated models after the GFC. The forecasting errors for short and medium yields in this period are relatively small in absolute terms, thus, it is highly likely that a poor relative forecasting performance is not picked up by commonly used global forecast evaluation measures such as the RMSE. Not considering the unique behaviour of short and medium yields in this time period may distort future results and interpretations.
Finally, it is important to develop mitigating measures to improve the relative forecasting accuracy in periods of flat, non-volatile interest rates. As a potential approach we identify forecast combination strategies. Simply equally combining all factor models already notably improves the inferior performance relative to the random walk. Combining two variations of Nelson-Siegel and PCA model, an AR(1) model directly applied on yield levels and the random walk, with model weights based on their recent forecasting performance significantly improves the forecasting accuracy, albeit the combination scheme is still not able to consistently beat the random walk. We also observe that it is typically different models that will provide the most appropriate forecasts through time. In particular during the transition from a more volatile behavior of interest rates to the current low yield environment with only minor fluctuations, weights allocated to the individual models change dramatically, with the random walk dominating the other models during the post GFC period. Overall, the results show that combining forecasts has the potential to significantly improve the forecasting accuracy especially for a time period where individual models perform poorly.
Our results also point towards the benefits of using adaptive forecasting techniques or regime switching models to predict the yield curve in different economic environment as they have recently been suggested by Xiang and Zhu (2013); Chen and Niu (2014). Such models may be more suitable to identify different phases of interest rate and yield curve behavior and may capture the change between volatile or quiet regimes also in their forecasts. Recent results, see, e.g., Koopman and van der Wel (2011); Exterkate et al.

(2013) have also shown that including macroeconomic variables can significantly improve the forecasting performance of yield curve models. It is thus worthwhile to investigate whether including macroeconomic variables may also enable the selected factor models to adjust quicker to the new interest rate environment and improve the forecasting accuracy. We leave these tasks for future research.

References

Ang, A., Bekaert, G., 2002. Regime Switches in Interest Rates. Journal of Business & Economic Statistics 20 (2), 163–182.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50 (4), 745–787.

Bianchetti, M., 2010. Two curves, one price. Risk August, 74–80.

Bikbov, R., Chernov, M., 2010. No-arbitrage macroeconomic determinants of the yield curve. Journal of Econometrics 159 (1), 166–182.

Blaskowitz, O., Herwartz, H., 2009. Adaptive forecasting of the EURIBOR swap term structure. Journal of Forecasting 28 (7), 575–594.

Carriero, A., Kapetanios, G., Marcellino, M., 2012. Forecasting government bond yields with large Bayesian vector autoregressions. Journal of Banking & Finance 36 (7), 2026–2047.

Chen, Y., Gwati, R., 2013. FX Options and Excess Returns: A Multi Moment Term Structure Model of Exchange Rate Dynamics.

Chen, Y., Niu, L., 2014. Adaptive dynamic Nelson–Siegel term structure model with applications. Journal of Econometrics 180 (1), 98–115.

Christensen, J. H., Diebold, F. X., Rudebusch, G. D., 2011. The affine arbitrage-free class of Nelson-Siegel term structure models. Journal of Econometrics 164 (1), 4–20.

Coroneo, L., Nyholm, K., Vidova-Koleva, R., 2011. How arbitrage-free is the Nelson-Siegel model? Journal of Empirical Finance 18 (3), 393–407.

Cox, J. C., Ingersoll, J. E., Ross, S. A., 1981. A Re-Examination of Traditional Hypotheses about the Term Structure of Interest Rates. The Journal of Finance 36 (4), 769.

Cox, J. C., Ingersoll, J. E., Ross, S. A., 1985. A Theory of the Term Structure of Interest Rates. Econometrica 53 (2), 385–407.

Dai, Q., Singleton, K. J., 2000. Specification Analysis of Affine Term Structure Models. Journal of Finance 55 (5), 1943–1978.

Dewachter, H., Lyrio, M., 2006. Macro Factors and the Term Structure of Interest Rates. Journal of Money, Credit and Banking 38 (1), 119–140.

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130 (2), 337–364.

Diebold, F. X., Li, C., Yue, V. Z., 2008. Global yield curve dynamics and interactions: A dynamic Nelson–Siegel approach. Journal of Econometrics 146 (2), 351–363.

Diebold, F. X., Mariano, R. S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13 (3), 253–263.

Diebold, F. X., Rudebusch, G. D., Aruoba, S. B., 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of Econometrics 131 (1-2), 309–338.

Duffee, G. R., 2002. Term Premia and Interest Rate Forecasts in Affine Models. The Journal of Finance 57 (1), 405–443.

Duffee, G. R., 2011. Forecasting with the term structure: The role of no-arbitrage restrictions.
URL `http://krieger2.jhu.edu/economics/wp-content/uploads/pdf/papers/wp576.pdf`

Duffee, G. R., 2012. Forecasting interest rates.

Duffie, D., Kan, R., 1996. A Yield-Factor Model Of Interest Rates. Mathematical Finance 6 (4), 379–406.

Exterkate, P., van Dijk, D., Heij, C., Groenen, P. J. F., 2013. Forecasting the Yield Curve in a Data-Rich Environment Using the Factor-Augmented Nelson-Siegel Model. Journal of Forecasting 32 (3), 193–214.

Fama, E. F., Bliss, R. R., 1987. The Information in Long-Maturity Forward Rates. The American Economic Review 77 (4), p 680–692.

Favero, C. A., Niu, L., Sala, L., 2012. Term Structure Forecasting: No-Arbitrage Restrictions versus Large Information Set. Journal of Forecasting 31 (2), 124–156.

Giacomini, R., Rossi, B., 2010. Forecast comparisons in unstable environments. Journal of Applied Econometrics 25 (4), 595–620.

Guidolin, M., Timmermann, A., 2009. Forecasts of US short-term interest rates: A flexible forecast combination approach. Journal of Econometrics 150 (2), 297–311.

Gürkaynak, R. S., Sack, B. P., Wright, J. H., 2007. The U.S. Treasury yield curve: 1961 to the present. Journal of Monetary Economics 54 (8), 2291–2304.

Hautsch, N., Yang, F., 2012. Bayesian inference in a Stochastic Volatility Nelson–Siegel model. Computational Statistics & Data Analysis 56 (11), 3774–3792.

Hull, J., White, A., 1990. Pricing Interest-Rate-Derivative Securities. Review of Financial Studies 3 (4), 573–592.

Koopman, S. J., van der Wel, M., 2011. Forecasting the U.S. Term Structure of Interest Rates using a Macroeconomic Smooth Dynamic Factor Model.

Koopman, S. J., van der Wel, M., 2013. Forecasting the US term structure of interest rates using a macroeconomic smooth dynamic factor model. International Journal of Forecasting 29 (4), 676–694.

Laurini, M. P., Hotta, L. K., 2014. Forecasting the Term Structure of Interest Rates Using Integrated Nested Laplace Approximations. Journal of Forecasting 33 (3), 214–230.

Leite, A. L., Filho, R. B. P. G., Vicente, J., 2010. Forecasting the yield curve: A statistical model with market survey data. International Review of Financial Analysis 19 (2), 108–112.

Litterman, R., Scheinkman, J., 1991. Common Factors Affecting Bond Returns. Journal of Fixed Income 1 (1), 54–61.

Martins, L. F., Perron, P., 2012. Forecast Comparisons in Unstable Environments: Comments and Improvements.

Moench, E., 2008. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. Journal of Econometrics 146 (1), 26–43.

Nelson, C. R., Siegel, A. F., 1987. Parsimonious Modeling of Yield Curves. The Journal of Business 60 (4), 473.

Pooter, M. D., Ravazzolo, F., van Dijk, D., 2010. Term structure forecasting using macro factors and forecast combination.

Reisman, H., Zohar, G., 2004. Short-Term Predictability of the Term Structure. Journal of Fixed Income 14 (3), 7–14.

Rudebusch, G. D., Wu, T., 2008. A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy*. The Economic Journal 118 (530), 906–926.

Svensson, L. E. O., 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994.

Timmermann, A., 2006. Forecast Combinations. In: Elliott, G., Granger, C. W. J., Timmermann, A. (Eds.), Handbook of economic forecasting. Vol. 1 of Handbook of Economic Forecasting. Elsevier North-Holland, Amsterdam and and Boston, pp. 135–196.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5 (2), 177–188.

Walker, S., McCormick, L., 2014. Unstoppable 100 Trillion Bond Market Renders Models Useless.
URL
http://www.bloomberg.com/news/2014-06-01/the-unstoppable-100-trillion-bond-market-renders-models-useless.html

Xiang, J., Zhu, X., 2013. A Regime-Switching Nelson-Siegel Term Structure Model and Interest Rate Forecasts. Journal of Financial Econometrics 11 (3), 522–555.

Yu, W.-C., Zivot, E., 2011. Forecasting the term structures of Treasury and corporate yields using dynamic Nelson-Siegel models. International Journal of Forecasting 27 (2), 579–591.

A US Yields - Diebold-Mariano statistics

		3m	6m	12m	2y	3y	5y	7y	10y
h=1	NSAR	4.97*	**-0.76**	**-1.45**	1.82	2.57*	3.77*	1.54	0.99
	NSVAR	1.84	**-1.11**	0.80	2.70*	1.66	3.20*	0.98	1.25
	PCAAR	2.13*	**-1.06**	0.23	4.20*	2.13*	1.24	0.47	0.12
	PCAVAR	0.00	**-1.51**	0.23	2.30*	1.42	1.48	1.06	**-0.11**
	AR1	1.93	1.80	1.49	1.45	1.23	0.85	0.71	0.62
h=6	NSAR	1.74	1.26	0.99	1.94	2.39*	2.15*	1.31	0.78
	NSVAR	**-0.10**	**-0.39**	**-0.32**	0.93	1.01	1.06	0.29	**-0.60**
	PCAAR	0.30	0.27	0.45	1.66	1.81	1.06	**-0.47**	**-2.17"**
	PCAVAR	**-0.24**	0.07	0.53	1.52	1.00	0.57	0.25	**-0.58**
	AR1	0.88	0.80	0.66	0.39	0.53	0.94	0.93	0.78
h=12	NSAR	1.08	0.91	1.01	1.64	1.94	2.15*	1.89	1.93
	NSVAR	**-0.61**	**-0.68**	**-0.36**	1.43	1.45	1.52	0.97	0.50
	PCAAR	**-0.48**	**-0.31**	**-0.04**	0.82	1.42	1.42	0.35	**-1.74**
	PCAVAR	**-0.54**	**-0.11**	0.47	2.10*	1.71	1.35	0.89	0.37
	AR1	0.63	0.54	0.38	0.08	0.49	1.30	1.52	1.53

Table 6. Diebold-Mariano forecast accuracy test-statistics of all investigated models against the random walk for US yields. We report the results of the period from **2004:01 to 2013:12** for one-month, six-months and twelve-months forecast horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. Note that **negative values indicate superiority of the investigated models against the the random walk**. (") denotes significance of the outperformance relative to the asymptotic null distribution at the 5% or smaller level. (*) denotes significance of the inferior performance against the random walk relative to the asymptotic null distribution at the 5% or smaller level. See section 4 for a description of the selected models.

		3m	6m	12m	2y	3y	5y	7y	10y
h=1	NSAR	2.57*	**-2.07"**	**-1.68**	1.54	1.24	1.92	1.05	0.48
	NSVAR	0.85	**-1.78**	**-0.79**	1.30	1.24	2.01*	0.89	0.15
	PCAAR	1.85	**-1.23**	**-0.12**	2.81*	1.38	1.03	0.80	0.39
	PCAVAR	**-0.35**	**-2.24"**	**-0.43**	1.56	1.11	1.02	1.20	0.82
	AR1	1.67	1.55	1.35	1.28	1.08	0.81	0.99	1.03
h=6	NSAR	0.19	**-0.77**	**-1.31**	**-0.05**	0.15	0.83	1.05	0.97
	NSVAR	**-1.07**	**-0.98**	**-0.64**	0.80	0.74	0.99	1.00	0.88
	PCAAR	**-0.33**	**-0.15**	0.05	0.50	0.15	0.19	0.03	**-0.91**
	PCAVAR	**-1.12**	**-0.43**	0.27	1.43	0.98	0.82	0.99	0.72
	AR1	0.86	0.83	0.69	0.03	**-0.67**	**-0.69**	0.37	0.56
h=12	NSAR	**-0.22**	**-0.75**	**-0.75**	**-0.09**	0.06	0.41	0.64	1.05
	NSVAR	**-1.80**	**-1.29**	**-0.84**	0.80	0.71	0.75	0.72	0.85
	PCAAR	**-1.24**	**-0.86**	**-0.58**	**-0.29**	**-0.45**	**-0.19**	0.01	**-0.21**
	PCAVAR	**-1.12**	**-0.46**	0.16	1.62	1.12	0.81	0.74	0.77
	AR1	0.64	0.57	0.40	**-0.32**	**-1.65**	**-0.77**	0.19	0.60

Table 7. Diebold-Mariano forecast accuracy test-statistics of the random walk against all selected models and the AR(1) model against all selected models for US yields. We report the results of the sub-sample period **2004:01-2009:12** for one-month, six-months and twelve-months forecast horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. Note that **negative values indicate superiority of the random walk**. (") denotes significance of the outperformance relative to the asymptotic null distribution at the 5% or smaller level. (*) denotes significance of the inferior performance against the random walk relative to the asymptotic null distribution at the 5% or smaller level. See section 4 for a description of the selected models.

		3m	6m	12m	2y	3y	5y	7y	10y
h=1	NSAR	11.53*	6.46*	1.43	1.15	3.32*	3.33*	1.14	0.86
	NSVAR	6.11*	4.81*	6.06*	2.95*	2.08*	2.51*	0.54	1.24
	PCAAR	4.51*	1.85	2.60*	4.04*	2.25*	0.78	**-0.25**	**-0.17**
	PCAVAR	4.02*	5.06*	3.51*	2.31*	1.47	1.10	0.18	**-0.77**
	AR1	1.51	1.20	0.70	1.08	0.65	0.42	0.11	0.02
h=6	NSAR	6.65*	6.67*	6.38*	7.89*	4.51*	1.90	0.85	0.25
	NSVAR	4.81*	3.31*	2.70*	**-0.43**	0.27	0.22	**-0.37**	**-1.49**
	PCAAR	3.12*	2.83*	3.35*	3.99*	2.59*	0.71	**-0.85**	**-2.37"**
	PCAVAR	2.78*	3.01*	3.26*	0.27	**-0.18**	**-0.27**	-0.47	**-1.13**
	AR1	1.90	1.21	0.44	1.68	1.49	0.86	0.61	0.39
h=12	NSAR	17.73*	0.00	0.00	0.00	8.85*	3.62*	2.09*	1.71
	NSVAR	0.00	0.00	9.92*	1.37	4.68*	1.42	0.16	0.00
	PCAAR	2.50*	2.34*	2.38*	3.13*	3.45*	1.64	0.00	0.00
	PCAVAR	4.18*	6.90*	7.11*	1.62	0.00	1.19	**-0.05**	0.00
	AR1	3.94*	0.00	0.50	1.84	2.10*	2.08*	1.63	1.25

Table 8. Diebold-Mariano forecast accuracy test-statistics of the random walk against all selected models and the AR(1) model against all selected models for US yields. We report the results of the sub-sample period **2009:01-2013:12** for one-month, six-months and twelve-months forecast horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. Note that **negative values indicate superiority of the random walk**. (") denotes significance of the outperformance relative to the asymptotic null distribution at the 5% or smaller level. (*) denotes significance of the inferior performance against the random walk relative to the asymptotic null distribution at the 5% or smaller level. See section 4 for a description of the selected models.

B US Yields Forecast Combination - Diebold Mariano statistics

		3m	6m	12m	2y	3y	5y	7y	10y
	CFEWfactor	1.56	**-2.02”**	**-0.51**	2.76*	1.47	2.29*	0.91	0.35
	CFPWfactor	0.94	**-2.23”**	**-0.85**	2.31*	1.28	1.98*	0.84	0.19
h=1	CFEWfar1	2.12*	**-1.57**	**-0.99**	3.41*	1.48	1.72	0.63	**-0.15**
	CFPWfar1	1.06	**-1.89**	**-1.29**	2.67*	1.20	1.35	0.53	**-0.42**
	CFEWfar1RW	1.79	**-1.75**	**-1.21**	3.16*	1.30	1.29	0.31	**-0.31**
	CFPWfar1RW	0.50	**-2.10**	**-1.55**	2.45*	0.99	0.92	0.17	**-0.61**
	CFEWfactor	**-0.41**	**-0.76**	**-0.79**	1.14	1.30	1.06	0.35	**-0.67**
	CFPWfactor	**-0.85**	**-1.18**	**-1.36**	0.18	0.47	0.79	0.20	**-0.85**
h=6	CFEWfar1	0.34	**-0.03**	**-0.08**	1.28	1.40	1.19	0.59	**-0.03**
	CFPWfar1	**-0.70**	**-0.96**	**-1.16**	0.00	0.39	0.98	0.51	**-0.22**
	CFEWfar1RW	0.19	**-0.16**	**-0.21**	1.13	1.16	0.89	0.31	**-0.29**
	CFPWfar1RW	**-0.91**	**-1.09**	**-1.28**	**-0.30**	**-0.10**	0.56	0.17	**-0.53**
	CFEWfactor	**-0.84**	**-0.93**	**-0.51**	1.35	1.54	1.63	1.17	0.80
	CFPWfactor	**-1.21**	**-1.33**	**-1.10**	0.30	0.62	1.08	0.89	0.31
h=12	CFEWfar1	**-0.24**	**-0.33**	**-0.03**	0.98	1.29	1.54	1.36	1.27
	CFPWfar1	**-1.02**	**-1.10**	**-0.95**	**-0.04**	0.41	1.13	1.20	0.97
	CFEWfar1RW	**-0.36**	**-0.44**	**-0.16**	0.86	1.16	1.36	1.12	0.97
	CFPWfar1RW	**-1.11**	**-1.18**	**-1.04**	**-0.28**	**0.00**	0.60	0.80	0.53

Table 9. Diebold-Mariano forecast accuracy test-statistics of the forecast combination strategies against the random walk for US yields. We report the results of the forecasting period **2004:01-2013:12** for one-month, six-months and twelve-months forecast horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. Note that **negative values indicate superiority of the investigated models against the the random walk**. (”) denotes significance of the outperformance relative to the asymptotic null distribution at the 5% or smaller level. (*) denotes significance of the inferior performance against the random walk relative to the asymptotic null distribution at the 5% or smaller level. See section 6 for a description of the selected combination strategies.

		3m	6m	12m	2y	3y	5y	7y	10y
h=1	CFEWfactor	1.10	**-2.22”**	**-0.97**	1.93	1.00	1.39	0.97	0.32
	CFPWfactor	0.81	**-2.35”**	**-1.12**	1.57	0.90	1.26	0.95	0.17
	CFEWfar1	1.50	**-1.78**	**-1.12**	2.88*	0.94	1.12	0.90	0.36
	CFPWfar1	0.93	**-2.00”**	**-1.38**	2.27*	0.75	0.94	0.86	0.04
	CFEWfar1RW	1.31	**-1.91**	**-1.28**	2.78*	0.86	0.96	0.77	0.31
	CFPWfar1RW	0.42	**-2.20**	**-1.59**	2.19*	0.63	0.73		
h=6	CFEWfactor	**-1.52**	**-1.39**	**-1.17**	0.33	0.19	0.58	0.76	0.44
	CFPWfactor	**-1.92**	**-1.86**	**-1.97”**	**-0.58**	**-0.30**	0.41	0.65	0.24
	CFEWfar1	**-0.63**	**-0.82**	**-0.80**	**-0.42**	**-0.83**	0.12	0.45	0.19
	CFPWfar1	**-1.37**	**-1.72**	**-2.00”**	**-1.63**	**-1.55**	**-0.18**	0.40	0.08
	CFEWfar1RW	**-0.69**	**-0.88**	**-0.85**	**-0.52**	**-1.10**	**-0.06**	0.26	**-0.01**
	CFPWfar1RW	**-1.50**	**-1.77**	**-2.04**	**-1.73**	**-1.76**	**-0.54**	0.12	**-0.17**
h=12	CFEWfactor	**-5.33”**	**-2.73”**	**-2.58”**	0.02	0.23	0.44	0.53	0.62
	CFPWfactor	**-4.08”**	**-3.39”**	**-3.31”**	**-0.60**	**-0.22**	0.19	0.45	0.52
	CFEWfar1	**-4.27”**	**-1.89**	**-1.63**	**-1.42**	**-0.64**	**-0.03**	0.27	0.45
	CFPWfar1	**-2.85”**	**-2.73”**	**-2.59”**	**-1.73**	**-1.06**	**-0.33**	0.20	0.30
	CFEWfar1RW	**-4.30**	**-1.94**	**-1.68**	**-1.47**	**-0.69**	**-0.10**	0.17	0.28
	CFPWfar1RW	**-2.66**	**-2.67**	**-2.53**	**-1.68**	**-1.13**	**-0.53**	0.02	0.14

Table 10. Diebold-Mariano forecast accuracy test-statistics of the forecast combination strategies against the random walk for US yields. We report the results of the sub-sample period **2004:01 to 2009:12** for one-month, six-months and twelve-months forecast horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. Note that **negative values indicate superiority of the investigated models against the the random walk**. (”) denotes significance of the outperformance relative to the asymptotic null distribution at the 5% or smaller level. (*) denotes significance of the inferior performance against the random walk relative to the asymptotic null distribution at the 5% or smaller level. See section 6 for a description of the selected combination strategies.

		3m	6m	12m	2y	3y	5y	7y	10y
h=1	CFEWfactor	4.84*	2.86*	3.14*	2.54*	1.86	1.84	0.30	0.21
	CFPWfactor	1.54	0.39	1.77	1.95	1.29	1.48	0.16	0.09
	CFEWfar1	5.72*	2.19*	1.00	1.86	1.61	1.33	0.03	**-0.42**
	CFPWfar1	0.18	**-0.14**	0.04	0.97	1.10	0.89	**-0.28**	**-0.56**
	CFEWfar1RW	4.65*	1.52	0.51	1.51	1.33	0.91	**-0.26**	**-0.57**
	CFPWfar1RW	**-0.58**	**-0.59**	**-0.57**	0.60	0.78	0.46	**-0.67**	**-0.73**
h=6	CFEWfactor	2.32*	2.23*	2.19*	4.04*	2.14*	0.54	**-0.17**	**-1.17**
	CFPWfactor	2.56*	2.24*	2.42*	**-0.47**	0.47	0.13	**-0.37**	**-1.35**
	CFEWfar1	3.63*	2.89*	2.65*	2.82*	2.32*	0.95	0.23	**-0.35**
	CFPWfar1	0.94	0.54	1.04	2.00*	1.78	0.77	0.04	**-0.55**
	CFEWfar1RW	3.23*	2.59*	2.59*	2.64*	1.99*	0.67	0.02	-0.55
	CFPWfar1RW	**-0.39**	**-0.73**	**-0.05**	1.76	1.34	0.36	**-0.27**	**-0.81**
h=12	CFEWfactor	2.26*	2.14*	2.22*	4.18*	6.14*	2.00*	0.74	**-0.15**
	CFPWfactor	3.02*	3.44*	4.21*	2.27*	0.00	1.32	0.24	**-2.85"**
	CFEWfar1	4.40*	3.18*	2.55*	2.91*	3.07*	2.15*	1.28	0.93
	CFPWfar1	1.40	1.39	0.03	2.13*	2.46*	1.86	0.98	0.50
	CFEWfar1RW	4.42*	3.08*	2.48*	2.78*	2.79*	1.81	1.01	0.63
	CFPWfar1RW	**-3.09**	**-2.47**	**-1.60**	2.12*	2.37*	1.48	0.52	0.00

Table 11. Diebold-Mariano forecast accuracy test-statistics of the forecast combination strategies against the random walk for US yields. We report the results of the sub-sample period **2009:01 to 2013:12** for one-month, six-months and twelve-months forecast horizons and three-months, six-months, twelve-months, two-year, three-year, five-year, seven-year and ten-year maturities. Note that **negative values indicate superiority of the investigated models against the the random walk**. (") denotes significance of the outperformance relative to the asymptotic null distribution at the 5% or smaller level. (*) denotes significance of the inferior performance against the random walk relative to the asymptotic null distribution at the 5% or smaller level. See section 6 for a description of the selected combination strategies.