

2010 | 01

Working Paper

Research Department

Term structure forecasting using macro factors and forecast combination

By Michiel de Pooter, Francesco Ravazzolo and Dick van Dijk

Working papers fra Norges Bank, fra 1992/1 til 2009/2 kan bestilles over e-post.

servicesenter@norges-bank.no

eller ved henvendelse til:

Norges Bank, Abonnementservice

Postboks 1179 Sentrum

0107 Oslo

Telefon 22 31 63 83, Telefaks 22 41 31 05

Fra 1999 og fremover er publikasjonene tilgjengelig på www.norges-bank.no

Working papers inneholder forskningsarbeider og utredninger som vanligvis ikke har fått sin endelige form. Hensikten er blant annet at forfatteren kan motta kommentarer fra kolleger og andre interesserte. Synspunkter og konklusjoner i arbeidene står for forfatterens regning.

Working papers from Norges Bank, from 1992/1 to 2009/2 can be ordered by e-mail:

servicesenter@norges-bank.no

or from Norges Bank, Subscription service

P.O.Box. 1179 Sentrum

N-0107Oslo, Norway.

Tel. +47 22 31 63 83, Fax. +47 22 41 31 05

Working papers from 1999 onwards are available on www.norges-bank.no

Norges Bank's working papers present research projects and reports (not usually in their final form) and are intended inter alia to enable the author to benefit from the comments of colleagues and other interested parties. Views and conclusions expressed in working papers are the responsibility of the authors alone.

ISSN 1502-8143 (online)

ISBN 978-82-7553-539-7 (online)

Term Structure Forecasting Using Macro Factors and Forecast Combination*

Michiel de Pooter[†]

Federal Reserve
Board

Francesco Ravazzolo

Norges Bank

Dick van Dijk

Erasmus University
Rotterdam

Abstract

We examine the importance of incorporating macroeconomic information and, in particular, accounting for model uncertainty when forecasting the term structure of U.S. interest rates. We start off by analyzing and comparing the forecast performance of several individual term structure models. Our results confirm and extend results found in previous literature that adding macroeconomic information, through factors extracted from a large number of individual series, tends to improve interest rate forecasts. We then show, however, that the predictive power of individual models varies over time significantly. Models with macro factors are the more accurate in and around recession periods. Models without macro factors do particularly well in low-volatility subperiods such as the late 1990s. We demonstrate that this problem of model uncertainty can be mitigated by combining individual model forecasts. Combining forecasts leads to encouraging gains in predictability, especially for longer-dated maturities, and importantly, these gains are consistent over time.

Keywords: Term structure of interest rates, Nelson-Siegel model, Affine term structure model, macro factors, forecast combination, Model Confidence Set

JEL classification: C5, C11, C32, E43, E47

*We thank Torben Andersen, Martin Martens, Dagfinn Rime, and Daniel Thornton for helpful discussions and for providing detailed comments, as well as seminar participants at the Catholic University Leuven, Erasmus University Rotterdam, Federal Reserve Bank of New York, Federal Reserve Board, Norges Bank, the 2008 Infinity Conference, and the 27th International Symposium on Forecasting. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other employee of the Federal Reserve System, nor do they reflect the views of Norges Bank (the Central Bank of Norway). This paper is best viewed in color. An earlier draft of this paper circulated under the name “Predicting the Term Structure of Interest Rates: Incorporating Parameter Uncertainty, Model Uncertainty and Macroeconomic Information” which is available as a Tinbergen Institute Discussion Paper (07-028/4).

[†]Corresponding author; De Pooter is a staff economist in the Division of International Finance, Board of Governors of the Federal Reserve System, Washington, D.C. 20551, U.S.A., Tel.: (202) 452-2264, fax: (202) 452-6424. *E-mail addresses:* michiel.d.depooter@frb.gov (M. De Pooter), francesco.ravazzolo@norges-bank.no (F. Ravazzolo), djvandijk@ese.eur.nl (D. van Dijk).

1 Introduction

Modelling and forecasting the term structure of interest rates is by no means an easy endeavor. Since long yields are risk-adjusted averages of expected future short rates, yields of different maturities are intimately related and therefore move together, in the cross-section as well as over time. At the same time, long and short maturities tend to react quite differently to shocks hitting the economy. Furthermore, monetary policy authorities such as the Federal Reserve are actively targeting the short end of the yield curve to achieve their macroeconomic goals. In general, many forces are at work at moving interest rates. Identifying these forces and understanding their impact on yields, is therefore of crucial importance.

In recent years, significant progress has been made in modelling the term structure of interest rates, which has come about mainly through the development of no-arbitrage factor models. The literature on these so-called affine term structure models was kick-started by seminal papers of Vasicek (1977) and Cox, Ingersoll, and Ross (1985), characterized by Duffie and Kan (1996) and classified by Dai and Singleton (2000). A survey of issues involving the specification and estimation of affine models set in continuous time is Piazzesi (2003). Discrete-time models are discussed in detail in Backus, Foresi, and Telmer (1998). Traditional affine models explain yield movements as being driven by a small number of (latent) factors that can be extracted from the panel of yields across time and across maturities, and impose cross-equation restrictions which are consistent with no-arbitrage. Affine models, provided they are properly specified, have been shown to accurately fit the term structure, see for example Dai and Singleton (2000). These models are rather silent, however, about the links between the (mainly) statistical yield factors and macroeconomic forces.

The current term structure literature is actively progressing to resolve this missing link. Recent studies have yielded interesting approaches for studying the joint behavior of interest rates and macroeconomic variables. One avenue that has been taken is to extend existing term structure models by adding in observed macroeconomic variables, and to study their interactions with the latent factors. A seminal contribution to this strand of the literature is Ang and Piazzesi (2003), who were the first to augment a standard three-factor affine model with macroeconomic variables. Studies such as Kim and Wright (2005), Dai and Philippon (2006), DeWachter and Lyrio (2006), Ang, Dong, and Piazzesi (2007), and Bikbov and Chernov (2008), among others, also incorporate various macroeconomic variables and study their explanatory power for yield movements. Studies that take a more structural approach include those by Wu (2005), Hordahl, Tristani, and Vestin (2006), and Rudebusch and Wu (2008), who all combine a model for the macro economy with an arbitrage-free specification for the term structure. Moving away from the realm of no-arbitrage interest rate models to that of more ad-hoc models, in particular the popular Nelson and Siegel (1987) model, studies such as Diebold, Rudebusch, and Aruoba (2006) and Mönch (2006) also show that

adding information which reflects the state of the economy is beneficial for explaining the level of interest rates.¹

Whereas fitting interest rate movements over time is already a strenuous task, accurately *forecasting* future interest rate levels is an even more difficult challenge. Yields of all maturities are close to being non-stationary, which makes it hard for any model to outperform the simple random walk no-change forecast. Several studies have documented that beating the random walk in terms of forecasting accuracy is indeed difficult, in particular for unrestricted yields-only vector autoregressive (VAR) and standard affine models, see Duffee (2002) and Ang and Piazzesi (2003). Recently, however, more favorable evidence for interest rate predictability has been reported. Duffee (2002) shows that more flexible affine specifications can beat the random walk. Diebold and Li (2006) and Christensen, Diebold, and Rudebusch (2009) show that dynamic Nelson-Siegel-style factor models forecast particularly well. Even more promising results are obtained with models that incorporate macroeconomic information. Ang and Piazzesi (2003) and Mönch (2008) report improved forecasts for U.S. Treasury yields at various horizons using affine models which have been augmented to include principal component-based macro factors. Hordahl, Tristani, and Vestin (2006) report similar improvements in predictability for German zero-coupon bond yields using inflation and industrial production. Ludvigson and Ng (2009) find that macro factors also help to forecast excess bond returns, indicating that macro factors contain predictive information that is not already contained in forward rates and yield spreads.

When examining the historical time series of U.S. interest rates we can easily identify subperiods across which yield curve dynamics appear to be quite different. This not only concerns characteristics such as the level and slope of the yield curve, but also the “stability” of the curve, that is, interest rate volatility. For example, the second half of the 1990s during which the yield curve was fairly stable, was followed by a strong and fast decline in interest rate levels in the early 2000s, accompanied by a pronounced widening of spreads when the Fed eased monetary policy in light of the burst of the dot-com bubble and the subsequent recession. Formal evidence of these kinds of different interest rate regimes is presented for example in Ang and Bekaert (2002).² It seems an overly daunting requirement for any individual model to be capable of consistently producing accurate forecasts under potentially very different interest rate regimes. In this paper, it is exactly this premise that we investigate for the term structure of U.S. interest rates. In order to do so we analyze a range of different models, from simple univariate autoregressive models to multivariate specifications with no-arbitrage restrictions, and we assess their forecasting performance over time.

¹Macro variables, however, mainly seem to help in capturing the dynamics of short and medium-term rates. Modelling long-term yields remains difficult. Dai and Philippon (2006) show that fiscal policy can account for some of the unexplained long rate dynamics whereas DeWachter and Lyrio (2006) show that long-run inflation expectations are important for modelling long-term bond yields.

²See also Bansal and Zhou (2002), Dai, Singleton, and Yang (2007), and the references contained therein.

We analyze each model in our model set with and without adding macroeconomic information to it. More specifically, we add macro factors, which we extract from a large set of individual macroeconomic variables. As noted above, several recent studies have shown that adding macroeconomic variables to term structure models helps to explain and forecast yield movements. Additionally, papers such as Ang and Piazzesi (2003), Mönch (2008) and Ludvigson and Ng (2009) document that using macro *factors*, extracted from a large panel of macro series, instead of *individual* series works well in affine models. We examine and extend this evidence by incorporating these types of macro diffusion indices also in the Nelson-Siegel model, as well as in simpler AR and VAR models. Our results show that adding macro factors does indeed improve the forecast accuracy of individual models. This only seems to be the case in particular interest rate regimes, however, and results vary across the term structure. As we demonstrate below, and which is part of the main message of this paper, we find that the predictive performance of individual models indeed varies over time considerably. Models that incorporate macroeconomic information are more accurate in subperiods with substantial uncertainty about the future path of interest rates. An example of a regime like this is in and around the 2001 recession. Models that do not include macroeconomic information do particularly well in subperiods where the term structure has a more stable pattern, or when the spread between long and short yields closes, as was the case in the second half of the 1990s for example.

The fact that different models forecast well in different subperiods confirms ex-post that different model specifications play a complementary role in approximating the unobserved data generating process of interest rates. Our results provide a strong incentive for examining forecast combination techniques as an alternative to believing in single models. We find that combining forecasts across all individual models, with and without macro factors, and after trimming out the worst performing models via Model Confidence Set tests as in Hansen, Lunde, and Nason (2003) gives accurate forecasts for short forecast horizons. Forecast combinations of just those models that include macro information, using a weighting method that is based on relative historical performance over a long sample, results in improved forecasts for long forecast horizons. Forecast accuracy in the latter case is particularly encouraging for longer-dated maturities, which traditionally have been difficult to forecast.

The remainder of the paper is organized as follows. In Section 2 we discuss the panel of U.S. Treasury yields we analyze in this study, and we provide details on the panel of macro series that we use in constructing our macro factors. We devote Section 3 to present the set of individual models in our model consideration set. In Section 4 we discuss forecast results of these individual models whereas in Section 5 we outline and analyze results of several forecast combination schemes. Finally, in Section 6 we conclude. The Appendices provide technical details on model inference and forecast evaluation criteria.

2 Data

2.1 Yield Data

Our term structure dataset consists of constant maturity, end-of-month continuously compounded yields on U.S. zero-coupon bonds. These have been constructed from average bid-ask price quotes on U.S. Treasuries from the CRSP government bond files. CRSP filters the available quotes by taking out illiquid bonds and bonds with option features. The remaining quotes are used to construct forward rates using the Fama and Bliss (1987) bootstrap method, as outlined in Bliss (1997). The forward rates are then averaged to construct constant maturity spot rates.³ Similar to Diebold and Li (2006) and Mönch (2008), our dataset consists of unsmoothed Fama-Bliss yields. These unsmoothed yields exactly price the underlying U.S. Treasury securities.

Throughout our analysis we use yields for $N = 13$ different maturities; $\tau = 1, 3$ and 6 months and 1, 2, ..., 10 years. We denote time- t yields by $y_t^{(\tau_i)}$ for $i = 1, \dots, N$. For the Nelson-Siegel models we follow Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006) by including additional maturities of 9, 15, 18, 21 and 30 months in order to increase the number of yield observations at the short end of the curve. Our sample period covers January 1970 till December 2003 for a total of 408 monthly observations. Similar to Duffee (2002) and Ang and Piazzesi (2003) we include data from well before the Volcker disinflation period, despite the reservations expressed in Rudebusch and Wu (2008) that it is likely that the pricing of interest rate risk and the relationship between yields and macroeconomic variables have changed during such a long time span. We do so for two reasons: (i) to have enough observations to identify the parameters of the models in our model consideration set with sufficient accuracy, as some models are highly parameterized, and (ii) to be able to assess forecasting performance over sufficiently long (sub-)periods with different yield curve characteristics.

The downside of using the Bliss dataset is that it stops at the end of 2003, well before the financial turmoil that started around July 2008 and which is obviously an interesting period during which to gauge the time-varying forecasting performance of various yield curve models. Two widely-used alternative datasets that contain more recent data are the Fama-Bliss CRSP dataset which is currently updated until the end of 2008, and the real-time dataset of Gürkanyak, Sack, and Wright (2007) (GSW) which is available from the Federal Reserve Board's website. The CRSP dataset only contains maturities up until five years, however, whereas one of our aims in this paper is to study model forecasting performance for longer-dated yields. The drawback of the GSW dataset is that it consists of smoothed fitted yields using the Svensson (1994) extension of the Nelson and Siegel (1987) model.

³We kindly thank Robert Bliss for providing us with the unsmoothed Fama-Bliss forward rates and the programs to construct the spot rates.

Since we include the two-step Nelson-Siegel specification of Diebold and Li (2006) as one of the models in our model consideration set (albeit that our first-round fitting step uses the original Nelson-Siegel model and not the Svensson extension as in GSW) we do not want to give this approach a potentially unfair advantage.

Figure 1(a) shows time-series plots for a subsample of the 13 maturities in our dataset whereas Table 1 reports summary statistics. The stylized facts common to yield curve data are clearly present: the sample average curve is upward sloping and concave, volatility is decreasing with maturity, autocorrelations are very high and increasing with maturity, and normality is rejected due to positive skewness and excess kurtosis. Correlations between yields of different maturities are high, especially for similar maturities. Even the maturities which are furthest apart (1 month and 10 years) still have a full-sample correlation as high as 86%.

2.2 Macroeconomic Data

Our macroeconomic dataset originates from Stock and Watson (2005) and consists of 116 series. Our macro dataset is the same as that of Ludvigson and Ng (2009). Contrary to Ludvigson and Ng (2009), however, we excluded all interest rate and interest rate spread-related series from the original 132 series in the dataset, discarding 16 series in total. We do include the federal funds rate as being an instrument for the stance of the Fed's monetary policy. The macro variables are classified in 15 categories: (1) output and income, (2) employment and hours, (3) retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) price indices, (14) average hourly earnings and (15) miscellaneous. Table 2 lists the series included in the macro dataset and the category they are classified in.

We transform the monthly recorded macro series, whenever necessary, to ensure stationarity by using log levels, annual differences or annual log differences. Column 2 of Table 2 lists the transformations. Outliers in each individual series are recursively replaced by the median value of the previous five observations, see Stock and Watson (2005) for details. We follow Ang and Piazzesi (2003), Diebold, Rudebusch, and Aruoba (2006), and Mönch (2008) and in our use of annual growth rates. Monthly growth rates series are very noisy and are therefore expected to add little information when added to the various term structure models.

We need to be careful about the timing of the macro series relative to the interest rate series to prevent the use of information that has not been released yet at the time when a forecast is made. This in order to make this a realistic pseudo real-time out-of-sample forecasting exercise. The interest rates in our dataset are recorded at the end of the month. Although macro figures tend to be released at the beginning or in the middle of the month,

they are typically released with a lag of one up to several several months. We accommodate for a potential look-ahead bias by lagging all macro series by one month, except for financial series; stock index variables, exchange rates and the federal funds rate, which are all monthly averages.⁴

Similar to Mönch (2008) and Ludvigson and Ng (2009), we extract a small number of common factors from our macro dataset. Mönch (2008), based on the work of Bernanke, Boivin, and Elias (2005), builds a no-arbitrage Factor-Augmented term structure model with four factors from a large panel of macroeconomic variables whereas Ludvigson and Ng (2009) use macro factors to predict excess bond returns. As in these papers, we apply principal component analysis to obtain macro *factors* from the full panel of macro *series*. Before extracting principal component factors, we first standardize all the series to have zero mean and unit variance, see Stock and Watson (2002a,b) for details. The use of common factors instead of individual macro series allows us to incorporate a much richer information set beyond that contained in often used variables such as CPI, PPI, employment, output gap or capacity utilization alone, while at the same time ensuring that the number of model parameters remains manageable.

For the full sample period, the first common macro factor explains 35% of the variation in the macro panel. The second and third factors explain an additional 19% and 8%, respectively, whereas the first 10 factors together explain an impressive 85%. Figure 2 shows the R^2 when regressing each individual macro series on each of first three factors separately. These types of regressions allows us to attach economic labels to the factors and to interpret them more as representing meaningful economic variables instead of simply as artifacts from a statistical procedure. The first factor closely resembles the series in the real output and employment categories (categories 1 and 2), as well as categories 3 through 8, and can therefore be labelled *business cycle* or *real activity* factor. The second factor loads mostly on inflation measures (category 13) which allows for the label of *inflation* factor. The third factor, although the correlations are much lower than for the first and second factor, is mostly related to money stock and reserves (category 12) and could thus be labelled a *monetary aggregates* or *money stock* factor. Figure 3 corroborates these interpretations graphically through time-series plots of the three macro factors together with industrial production (total), consumer price index (all items) and money stock (M1), respectively.

We have chosen to include the first three factors as exogenous explanatory variables in the various term structure models because, together, these factors explain over 60% of the

⁴Using contemporaneous information may exaggerate the benefits of using macroeconomic information when forecasting yields. Note, however, that we would only be able to fully mimic the information available to the econometrician at the time of making any forecast if we would use vintage data. Croushore (2006) discusses the use of vintage data and shows that data revisions can lead to an improvement in perceived forecastability. Here we use only revised final-vintage macroeconomic series, implying that this may affect our results as well.

variation in the macro panel.⁵ Given that we want to construct interest rate forecasts we also need to select a model to forecast the macro factors. We discuss this in more detail in Section 3.1.

3 Models

We assess the individual and combined forecasting performance of a range of models that are commonly used in the literature as well as by practitioners. Since previous studies have shown that parsimonious models often outperform more sophisticated models, we consider models with different levels of complexity. Our model set ranges from unrestricted linear specifications for yield levels (AR and VAR models), models that impose a parametric structure on factor loadings (the Nelson-Siegel class of models), to models that impose cross-sectional restrictions to rule out arbitrage opportunities (affine models). Our benchmark model throughout our forecasting exercise is the random walk model.

We could in principle consider an almost unlimited number of different models. For example, one can think of lots of different models resulting from including various (subsets of) individual macro variables, such as the models of Diebold, Rudebusch, and Aruoba (2006) and Hordahl, Tristani, and Vestin (2006). Although it is true that these models can be more economically meaningful than some of the models we examine, considering each and every one of these would blow up the number of models in our consideration set. To keep the number manageable, we therefore consider only a small but representable subset of models. Furthermore, we circumvent the decision of which individual macro variables to include by basically including all of them through our macro factor approach.

In this section we present the different models. We defer all specific details regarding inference and generating (multi-step ahead) forecasts to Appendix A.

3.1 Incorporating macro factors

The approach we use to incorporate the three macro factors is the following. Denote M_t as the (3×1) vector containing the time- t values of the macro factors. We add the factors to

⁵As a robustness check we also examined using additional factors, but the forecasting results were very similar. With fewer factors (one or two) we obtained worse results. Note that we made a somewhat ad hoc choice for the number of factors, based solely on how much of the variance each factor explains in the cross section of macro series. An alternative, and arguably better approach, would be to select the number, as well as which factors, by using information criteria or by selecting only factors that are judged to have predictive power for interest rates. Although certainly interesting, we leave this for future research. Ludvigson and Ng (2009) use such an approach to select their factors. One interesting difference resulting from their approach compared to ours is that they find that they need to include a stock market factor. In our sample, the 7th PCA factor is most related to stock market variables, but explains only 3% of the variance in the macro panel and hence does not make the cut to be included in our vector of macro factors that we incorporate in the models.

each term structure model, contemporaneously as well as lagged by one month to capture any delayed effects of macroeconomic news on the term structure.⁶ The exogenous explanatory macro information we add to the models is denoted by X_t , and is thus given by $X_t = (M_t' M_{t-1}')'$.

Our approach implies that when we forecast yields, we also need to model and forecast the macro factors. We tackle this issue by following Ang and Piazzesi (2003) in only allowing for a *unidirectional* link from macro variables to yields. Although this can be argued to be a restrictive assumption as it does not allow for a potentially rich *bidirectional* feedback, it enables us to model the time-series behavior of the macro factors separate from that of yields, which considerably facilitates estimation.⁷ Information criteria suggest modeling and forecasting M_t using a VAR model with three lags:

$$M_t = c + \Phi_1 M_{t-1} + \Phi_2 M_{t-2} + \Phi_3 M_{t-3} + \xi_t, \quad \varepsilon_t \sim \mathcal{N}(0, H) \quad (1)$$

where c is a (3×1) vector, Φ_i is a (3×3) matrix for $i = 1, \dots, 3$, and H is a (3×3) unrestricted covariance matrix. Forecasts of future factor values can be constructed by forward iteration of the estimated relationship in (1).

3.2 Interest Rate Models

Random walk

The first model that we consider is a random walk without drift for each individual maturity τ_i , $i = 1, \dots, N$,

$$y_t^{(\tau_i)} = y_{t-1}^{(\tau_i)} + \varepsilon_t^{(\tau_i)}, \quad \varepsilon_t^{(\tau_i)} \sim \mathcal{N}\left(0, \sigma^{(\tau_i)^2}\right) \quad (2)$$

In this model any h -step ahead forecast $\hat{y}_{T+h}^{(\tau_i)}$ is simply equal to the most recently observed value $y_T^{(\tau_i)}$. It is natural to consider this no-change model as the benchmark against which to judge the predictive power of other models, and we do so throughout the paper. Table 1 confirms that yields are indeed all but non-stationary as the reported first-order autocorrelation coefficients are all very close to unity. Duffee (2002), Ang and Piazzesi (2003), Diebold and Li (2006), and Mönch (2008) all show, using different models and different forecast periods, that beating the random walk in terms of forecasting performance is quite an arduous task. We denote the random walk model by the abbreviation **RW**.

⁶Note again that “contemporaneous” here means that we use financial series recorded at time t , whereas time $t - 1$ values are used for the remaining macro series, see Section 2.2 for further details.

⁷In a forecasting exercise using German zero-coupon yields, Hordahl, Tristani, and Vestin (2006) show that term-structure information helps little in forecasting macroeconomic variables (more specifically (i) inflation and (ii) the output gap) which provides an argument for forecasting macro variables outside of term structure models. The authors note, however, that this might be due to the fact that their proposed macroeconomic model has an imperfect ability to describe the joint dynamics of German macroeconomic variables. On the other hand, Diebold, Rudebusch, and Aruoba (2006) and Ang, Dong, and Piazzesi (2007) do allow for bi-directional effects between macro variables and latent yield factors but both studies find that the causality from macro variables to yields is much stronger than vice versa.

AR model

Although (unreported) results indicate that the null of a unit root for yield levels cannot be rejected statistically, the assumption of nonstationary yields is difficult to interpret from an economic point of view. Nonstationarity implies that interest rates can roam around freely and do not revert back to a long-term mean, something which contradicts the Federal Reserve’s monetary policy objective of moderate long-term interest rates. The second model that we consider therefore is a first-order univariate autoregressive model which allows for mean-reversion,

$$y_t^{(\tau_i)} = c^{(\tau_i)} + \phi^{(\tau_i)} y_{t-1}^{(\tau_i)} + \psi^{(\tau_i)'} X_t + \varepsilon_t^{(\tau_i)}, \quad \varepsilon_t^{(\tau_i)} \sim \mathcal{N}\left(0, \sigma^{(\tau_i)2}\right) \quad (3)$$

where $c^{(\tau_i)}$, $\phi^{(\tau_i)}$ and $\sigma^{(\tau_i)}$ are scalar parameters and $\psi^{(\tau_i)}$ is a (6×1) vector containing the coefficients on the macro factors. We construct forecasts both with and without macro factors by setting $\psi^{(\tau_i)} = 0$. We denote the yield-only model by **AR** and the model with macro factors by **AR-X**. For this and all other models we construct *iterated* h -step ahead forecasts. Another approach is to construct *direct* forecasts, by regressing $y_t^{(\tau_i)}$ directly on its h -month lagged value $y_{t-h}^{(\tau_i)}$ as in Diebold and Li (2006). For the state-space form of the Nelson-Siegel model and the affine model such an approach is, however, uncommon. For the sake of consistency, we therefore chose to use iterated forecasts for all the models. Whether iterated forecasts are more accurate than direct forecasts is still an ongoing debate, see for example the recent discussion in Marcellino, Stock, and Watson (2006). In the context of interest rate forecasting, Carriero, Kapetanios, and Marcellino (2009) find that for linear AR and VAR models the iterated approach produces better forecasts than the direct approach.

VAR model

Vector autoregressive (VAR) models allow for using the history of other maturities as additional information on top of any maturity’s own history. We use the following first-order VAR specification,⁸

$$Y_t = c + \Phi Y_{t-1} + \Psi X_t + H \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, I) \quad (4)$$

where Y_t contains the yields for all 13 maturities; $Y_t = [y_t^{(1m)}, \dots, y_t^{(10y)}]'$, c is a (13×1) vector, Φ a (13×13) matrix, Ψ a (13×6) matrix, and H is the (unrestricted) residual variance matrix containing $\frac{1}{2}N(N + 1) = 91$ free parameters. Our approach is similar in spirit to the VAR models used in Evans and Marshall (1998, 2007) and Ang and Piazzesi

⁸For both the AR and VAR models we examined the benefits of including more lags by analyzing AR(p) and VAR(p) models with $p = 2, \dots, 12$. We found that using multiple lags resulted in nearly identical forecasts compared to the AR(1) and VAR(1) models and these results are therefore not reported, nor are they included in the forecasting combination procedures in Sections 4 and 5.

(2003) in the sense that we impose exogeneity of macroeconomic variables with respect to yields.

A well-known drawback of using an unrestricted VAR model for yields is that forecasts can only be constructed for those maturities that are actually included in the model. Since we want to construct forecasts for thirteen maturities, this results in a substantial number of parameters that need to be estimated. In an attempt to mitigate estimation error and, consequently, to reduce the forecast error variance, we instead summarize the information contained in the explanatory vector Y_{t-1} by replacing it with a small number of common yield curve factors. Similar to Litterman and Scheinkman (1991) and many other studies, we find that the first 3 principal components explain almost all the variation in the cross section of yields (over 99% for the full sample). Accordingly, we replace Y_{t-1} in (4) with the (3×1) vector of yield factors F_{t-1} :

$$Y_t = c + \Phi F_{t-1} + \Psi X_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, H) \quad (5)$$

where Φ is now a (13×3) matrix. The VAR model without and with macroeconomic variables is denoted by **VAR** and **VAR-X**, respectively.

Nelson-Siegel model

Diebold and Li (2006) show that using the in essence static Nelson and Siegel (1987) model as a dynamic factor model generates highly accurate interest rate forecasts. The Nelson-Siegel model differs from the unrestricted VAR model in (5) in that it imposes a parametric structure on the factor loadings. The factor loadings Φ are specified as exponential functions of time to maturity and a single parameter λ . Following Diebold, Rudebusch, and Aruoba (2006), the state-space representation of the three-factor model, with a first-order autoregressive model for the dynamics of the state vector, is given by

$$y_t^{(\tau_i)} = \beta_{1,t} + \beta_{2,t} \left[\frac{1 - \exp(-\tau_i/\lambda)}{\tau_i/\lambda} \right] + \beta_{3,t} \left[\frac{1 - \exp(-\tau_i/\lambda)}{\tau_i/\lambda} - \exp(-\tau_i/\lambda) \right] + \varepsilon_t^{(\tau_i)} \quad (6)$$

$$\beta_t = a + \Gamma \beta_{t-1} + u_t \quad (7)$$

The state vector, $\beta_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$, contains the latent factors at time t which can be interpreted as level, slope and curvature factors, respectively (see Diebold and Li, 2006 for details). The parameter λ governs the exponential decay towards zero of the factor loadings on $\beta_{2,t}$ and $\beta_{3,t}$, a is a (3×1) vector of parameters, and Γ is a (3×3) parameter matrix. We assume that the measurement equation and state equation errors in (6) and (7) are normally distributed and mutually uncorrelated;

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0_{18 \times 1} \\ 0_{3 \times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) \quad (8)$$

where H is a diagonal (18×18) matrix and Q a full (3×3) matrix. We follow Diebold and Li (2006) by adding five maturities ($\tau = 9, 15, 18, 21$ and 30 months) to the short end of

the yield curve to estimate the Nelson-Siegel model in (6)-(8). To estimate the Nelson-Siegel model, we use two different estimation procedures: a two-step approach and a one-step approach.

The two-step approach is used in Diebold and Li (2006) and consists of first estimating the latent factors in β_t using the cross-section of yields for each month t , while fixing λ . Given the estimated time-series for the factors, the second step then consists of modeling the dynamics of the factors in (7) by fitting either a joint VAR(1) model, or by estimating separate AR(1) models, thereby assuming that both Γ and Q are diagonal. We denote these approaches by **NS2-VAR** and **NS2-AR**, respectively. The one-step approach follows from Diebold, Rudebusch, and Aruoba (2006) and involves jointly estimating (6)-(8) as a state space model using the Kalman filter. In this approach we assume that Γ and Q are both full matrices, while λ is now estimated alongside the other parameters. We denote the one-step approach by **NS1**.

Diebold, Rudebusch, and Aruoba (2006) show how to extend the Nelson-Siegel model to incorporate macroeconomic variables by adding these as observable factors to the state vector, and then writing the model in companion form:

$$y_t^{(\tau_i)} = \beta_{1,t} + \beta_{2,t} \left[\frac{1 - \exp(-\tau_i/\lambda)}{\tau_i/\lambda} \right] + \beta_{3,t} \left[\frac{1 - \exp(-\tau_i/\lambda)}{\tau_i/\lambda} - \exp(-\tau_i/\lambda) \right] + \varepsilon_t^{(\tau_i)} \quad (9)$$

$$f_t = a + \Gamma f_{t-1} + \eta_t \quad (10)$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0_{18 \times 1} \\ 0_{12 \times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) \quad (11)$$

The state vector now also contains observable factors; $f_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, M_t, M_{t-1}, M_{t-2})'$.⁹ The dimensions of a , Γ and Q are increased appropriately and η_t is now given by $\eta_t = (u'_t, \xi'_t, 0, \dots, 0)'$. We impose structure on Γ and Q to accommodate for the effects of lagged macro factors while maintaining the unidirectional causality from macro factors to yields only.¹⁰ In particular, the lower left (9×3) block of Γ consists of zeros whereas Q is block diagonal with a non-zero (3×3) block Q_β for the yield factors and a non-zero (3×3) block Q_M for the contemporaneous macro factors. All other blocks on the diagonal contain zeros only. The Nelson-Siegel model with macro factors can also again be estimated by using either a two-step approach with AR or VAR dynamics for the yield factors, which we denote by **NS2-AR-X** and **NS2-VAR-X**, respectively, or by using the one-step approach, which we denote by **NS1-X**. Another potential specification of the Nelson-Siegel model would be

⁹Note that because we model the observable macro factors in M_t with a VAR(3) model, we need to add both the first and second lag, M_{t-1} and M_{t-2} , respectively, to the state vector in order to write the state equations in companion form.

¹⁰The macro factors are prevented from entering the measurement equations directly by only allowing the factor loadings of β_t to be non-zero in (9). Diebold, Rudebusch, and Aruoba (2006) impose this restriction to maintain the assumption that three factors are sufficient to describe interest rate dynamics. We follow Diebold, Rudebusch, and Aruoba (2006) here because relaxing this assumption would result in a substantial number of additional parameters.

that of Christensen, Diebold, and Rudebusch (2009) who adjust the Nelson-Siegel model to make it consistent with arbitrage-free models (to be discussed in the next section). Although Christensen, Diebold, and Rudebusch (2009) show that the Arbitrage-Free Dynamic Nelson-Siegel (AFDNS) model forecasts well out-of-sample, Carriero, Kapetanios, and Marcellino (2009), using a longer forecasting sample, report that the performance of the AFDNS model is not that different from the two-step Nelson-Siegel model. Because our model set is already large as it is, we therefore chose not to include the AFDNS model in our model set.

Affine model

Models that impose no-arbitrage restrictions have been examined for their forecast accuracy in for example Duffee (2002), Ang and Piazzesi (2003) and Mönch (2008). The attractive property of the class of no-arbitrage models is that sound theoretical cross-sectional restrictions are imposed on factor loadings to rule out arbitrage opportunities. In this paper we consider a Gaussian-type discrete time affine no-arbitrage model, using a set-up similar to Ang and Piazzesi (2003). In particular, we assume that movements in the yield curve are driven by a vector of K underlying state variables, Z_t , which we assume follows a Gaussian VAR(1) process

$$Z_t = \mu + \Psi Z_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma \Sigma') \quad (12)$$

where Σ is a $(K \times K)$ lower triangular Choleski matrix, μ a $(K \times 1)$ parameter vector and Ψ a $(K \times K)$ parameter matrix.

The short interest rate is assumed to be an affine function of the factors

$$r_t = \delta_0 + \delta_1' Z_t \quad (13)$$

where δ_0 is a scalar and δ_1 a $(K \times 1)$ vector. We adopt a standard form for the pricing kernel, which is assumed to price all assets in the economy,

$$m_{t+1} = \exp\left(-r_t - \frac{1}{2} \lambda_t' \lambda_t - \lambda_t' u_{t+1}\right)$$

We specify market prices of risk to be time-varying and affine in the state variables

$$\lambda_t = \lambda_0 + \lambda_1 Z_t \quad (14)$$

with λ_0 a $(K \times 1)$ vector and λ_1 a $(K \times K)$ matrix. Risk premia are constant over time if λ_1 is equal to a zero matrix. When λ_0 is also equal to zero, risk premia are zero altogether.

Under the above assumptions it can be shown that bond prices are an exponentially-affine function of the state variables,

$$P_t^{(\tau)} = \exp[A^{(\tau)} + B^{(\tau)'} Z_t] \quad (15)$$

We can recursively determine the price of a τ -period bond using

$$P_t^{(\tau)} = \mathbb{E}_t[m_{t+1} P_{t+1}^{(\tau-1)}] \quad (16)$$

where the expectation is taken under the risk-neutral measure. Ang and Piazzesi (2003), among others, show that this gives the following recursive formulas for the bond pricing coefficients $A^{(\tau)}$ and $B^{(\tau)}$:

$$A^{(\tau+1)} = A^{(\tau)} + B^{(\tau)'}[\mu - \Sigma\lambda_0] + \frac{1}{2}B^{(\tau)'}\Sigma\Sigma'B^{(\tau)} - \delta_0 \quad (17)$$

$$B^{(\tau+1)'} = B^{(\tau)'}[\Psi - \Sigma\lambda_1] - \delta_1' \quad (18)$$

when starting from $A^{(0)} = 0$ and $B^{(0)} = 0$. If bond *prices* are exponentially affine in the state variables then *yields* are affine in the state variables since $P_t^{(\tau)} = \exp[-y_t^{(\tau)}\tau]$. Consequently, it follows that $y_t^{(\tau)} = a^{(\tau)} + b^{(\tau)'}Z_t$ with $a^{(\tau)} = -A^{(\tau)}/\tau$ and $b^{(\tau)} = -B^{(\tau)}/\tau$. To estimate the model we deviate from the popular Chen and Scott (1993) approach and instead assume that every yield is contaminated with measurement error in a state-space estimation set-up.

To summarize, we specify the following affine model

$$y_t^{(\tau_i)} = a^{(\tau_i)} + b^{(\tau_i)}Z_t + \varepsilon_t^{(\tau_i)} \quad (19)$$

$$Z_t = \mu + \Psi Z_{t-1} + u_t \quad (20)$$

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{13 \times 1} \\ 0_{3 \times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix}\right) \quad (21)$$

where H is assumed to be a diagonal matrix, $Q = \Sigma\Sigma'$, and $a^{(\tau_i)}$ and $b^{(\tau_i)}$ are the recursive yield equation functions. We assume Z_t to consist of $K = 3$ common factors. We denote this model by **ATSM**.

We extend the model to incorporate observable macroeconomic factors in a similar way as for the Nelson-Siegel model,

$$y_t^{(\tau_i)} = a^{(\tau_i)} + b^{(\tau_i)}f_t + \varepsilon_t^{(\tau_i)} \quad (22)$$

$$f_t = \mu + \Psi f_{t-1} + \eta_t \quad (23)$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{13 \times 1} \\ 0_{12 \times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix}\right) \quad (24)$$

with $f_t = (Z_t, M_t, M_{t-1}, M_{t-2})$. The state equation (23) is written in companion form and the dimensions of $a^{(\tau_i)}$, $b^{(\tau_i)}$, μ , Ψ and Q are again increased appropriately. As in the Nelson-Siegel model, Q is block diagonal with only two non-zero blocks, Q_Z and Q_M . Unlike in the Nelson-Siegel model, however, in the affine model yield movements are also directly related to current and past macro movements through the bond pricing coefficients. We do assume that the short rate and risk premia only depend on contemporaneous values of the macro factors, i.e. we set all coefficients in δ_0 , δ_1 , λ_0 and λ_1 associated with M_{t-1} and M_{t-2} equal to zero, similar to the ‘macro model’ in Ang and Piazzesi (2003). We denote the affine model with macroeconomic factors by **ATSM-X**.

We want to note two points here. First, our affine-with-macro model is a hybrid between the macro model of Ang and Piazzesi (2003) and the FAVAR model of Mönch (2008). Compared to Ang and Piazzesi (2003) we use macro factors that are based on many more macro

variables, whereas compared to Mönch (2008) we also incorporate latent yield factors. The yield factors are likely to improve the predictive ability of the model because the yield factors can better pick up high-frequency movements in yields (see also the discussion in Mönch, 2008). Second, we estimate the affine model using the Kalman Filter where we assume that every yield has measurement error. This implies that the factors in f_t are not simply a linear combination of yields so that the macro factors do truly add exogenous information to the model.

Adding macroeconomic variables or factors to affine models can cause estimation problems because it further increases the number of parameters in these already highly parameterized models.¹¹ To speed up as well as to facilitate the estimation procedure, we therefore use the two-step approach of Ang, Piazzesi, and Wei (2006) by making the latent yield factors observable. Contrary to Ang, Piazzesi, and Wei (2006), however, who directly use the observed short rate and the term-spread as measures of the level and slope of the yield curve, we use principal component analysis to extract common factors from the full set of yields. We use the first three factors as our observable state variables.

4 Forecasting

4.1 Forecast procedure

We divide our dataset into an initial estimation sample which covers the period 1970:1 - 1988:12 (228 observations) and a forecasting sample which is comprised of the remaining period 1989:1 - 2003:12 (180 observations). The first sixty months of the forecast period are used as a training sample to start up the forecast combinations discussed in Section 5. Consequently, we report forecast results for the sample 1994:1 - 2003:12 (120 observations).

We recursively estimate models using an expanding window, starting from the initial sample 1970:1 - 1988:12.¹² Given a set of parameter estimates, we construct point forecasts for four different horizons: $h = 1, 3, 6$ and 12 months ahead. As discussed in the previous section, for horizons beyond $h = 1$ month we compute iterated forecasts. To prevent data-snooping, we also recursively construct the macroeconomic factors (see Section 2.2), as well as the yield curve factors used in the VAR and the ATSM.

¹¹Contrary to the reduced-form affine model of Ang and Piazzesi (2003), Hordahl, Tristani, and Vestin (2006) use a structural affine model with macroeconomic variables in which the number of parameters can be kept down. They show that their model leads to better longer horizon interest rate forecasts than the Ang and Piazzesi (2003) model. These results indicate that instead of only imposing no-arbitrage restrictions, which is the case in affine models, imposing also *structural* equations seems to mitigate overparameterization.

¹²To address the Lucas Critique and to check the robustness of our results, we also repeated our analysis using a moving window of ten years. Although somewhat surprising perhaps, results were rather similar to the expanding window results which we discuss below.

4.2 Forecast evaluation

To evaluate out-of-sample forecasts we compute popular error metrics, per maturity and per forecast horizon. For a full sample evaluation we compute the Root Mean Squared Prediction Error (RMSPE). Similar to Hordahl, Tristani, and Vestin (2006) we also summarize the forecasting performance of each model across *all* maturities for a given forecast horizon by computing the Trace Root Mean Squared Prediction Error (TRMSPE), see Christoffersen and Diebold (1998) for details.

The drawback of using (T)RMSPE statistics is, however, that these are *single* statistics summarizing individual forecasting errors over an *entire* sample. Although often used, unfortunately they do not give any insight as to *where* in the sample models make their largest and smallest forecast errors. We therefore also graphically analyze the Cumulative Squared Prediction Errors (CSPE) used in Welch and Goyal (2008). These cumulative squared prediction error series clearly show in which months models outperform and in which months they underperform a given benchmark (here the random walk model). The model- m , time- T CSPE for a τ_i -month maturity is given by

$$\text{CSPE}_{m,T}(\tau_i) = \sum_{t=1}^T \left[\left(\widehat{y}_{t+h|t,\text{RW}}^{(\tau_i)} - y_{t+h}^{(\tau_i)} \right)^2 - \left(\widehat{y}_{t+h|t,m}^{(\tau_i)} - y_{t+h}^{(\tau_i)} \right)^2 \right] \quad (25)$$

where $y_{t+h}^{(\tau_i)}$ is the yield for a τ_i -month maturity observed at time $t+h$, while $\widehat{y}_{t+h|t,m}^{(\tau_i)}$ is its model- m forecast, made at time t . See Appendix B for further detailed formulas.

To test for statistically significant differences in forecasting accuracy between competing models we apply the Model Confidence Set (MCS) approach developed by Hansen, Lunde, and Nason (2003, 2005). Given a set of competing forecasting models, M_0 , the MCS procedure identifies the MCS $\widehat{M}_\alpha^* \subset M_0$, which is the set of models that contains the “best” forecasting model given a confidence level $1-\alpha$. Starting from the full set of models, $M = M_0$, and a vector of R forecasts, the MCS procedure repeatedly tests the null hypothesis of equal forecasting accuracy,

$$H_{0,M} : E[d_{ij,t}] = 0, \text{ for all } i, j \in M,$$

where $d_{ij,t} = L_{i,t} - L_{j,t}$ is the loss differential between models i and j in the set, with L being an appropriate loss function. The MCS procedure sequentially eliminates the worst performing models from M as long as the null is rejected. This procedure is repeated until the null is no longer rejected, in which case the surviving set is \widehat{M}_α^* . We follow Hansen, Lunde, and Nason (2003) by using their semi-quadratic statistic which gives the following t -statistics:

$$T_{SQ} \equiv \sum_{i,j \subset M} t_{ij}^2,$$

where $t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\text{var}(\bar{d}_{ij})}}$ for $i, j \subset M$ and $\bar{d}_{ij} = \frac{1}{R} \sum_{t=T}^{T+R-1} d_{ij,t}$. Similarly, we implement the

MCS procedure using the stationary block bootstrap of Politis and Romano (1994) with an average block length of 20 months and we the squared forecast error as loss function.

In the tables below we report results for confidence levels of $1 - \alpha = 90\%$ and $1 - \alpha = 75\%$. The test is performed independently for different maturities and forecast horizons.

4.3 Forecasting results: individual models

We start our discussion of the forecasting performance of individual models by considering the results in Panels A and B of Tables 3 to 6. The first row of each table reports the (T)RMSPE for the random walk model, whereas the remaining rows in Panels A and B are (T)RMSPEs *relative* to those of the random walk. Any number below one therefore indicates outperformance relative to the random walk, whereas any number larger than one signals underperformance. Two stars next to the RMSPE individual models indicates that a model belongs to the model set $\widehat{\mathcal{M}}_{0.25}^*$ according to the T_{SQ} test statistic, whereas one star is for when it belongs to the model set $\widehat{\mathcal{M}}_{0.10}^*$ instead. Figures 12 to 15 show time-series plots of the realized and predicted yields, both for individual models as well as for forecast combination methods (discussed in Section 5)

At first sight the results in Tables 3 to 6 are disappointing if we focus solely on the TRMSPE results in the first column of each table. There is not a single model that, across the board of maturities, consistently outperforms the random walk for all forecast horizons, as reflected by the relative TRMSPE statistics. In addition, when considering each horizon in isolation, still only a few models produce forecasts which are more accurate than simply repeating the last known value, and for those that do the improvements are often only marginal. The univariate autoregressive model augmented with macro factors gives the lowest TRMSPE for short horizons (1 and 3 months), whereas the VAR model with macro factors does so for longer horizons (6 and 12 months). More complex models such as the affine and Nelson-Siegel models perform poorly.

Focusing on specific maturities gives us more and different insights however. Predictability tends to be relatively high for short forecast horizons and short maturities as evident from the relative RMSPE statistics. For example, for the 1-month yield the majority of models outperform the random walk at both the 1-month and 3-month forecast horizon. Moreover, for both horizons the random walk is not in the final full-sample Model Confidence Set. For medium maturities, such as the 1-year and 2-year yield, the random walk is more difficult to beat, although the MCS tends to be smallest for these yields, consisting primarily of the random walk and the AR-X model. Although some models still provide RMSPE statistics below one for long maturities, only a few models, if any, are dropped from the final MCS. For example, for the 10-year yield all models end up in the MCS at the 3-month horizon.

For the 6-month and 12-month forecast horizons, using macroeconomic information seems to be a pre-requisite for obtaining at least some level of predictability. Among the macro-

augmented yield models, the VAR-X model outperforms the random walk most consistently across maturities, in particular for a 12-month horizon. Contrary to its results for shorter forecast horizons, the AR-X model is now accurate only for short maturities. Interestingly, the most accurate forecasting models for short maturities are the NS1-X and ATSM-X models. For medium and longer-dated maturities, imposing no-arbitrage restrictions on factor loadings does not help in forecasting yields. This result is consistent with Duffee (2009) who argues that no-arbitrage restrictions have no practical effect on forecast accuracy.

With the exception of one case - the ATSM for the 1-month yield for a 6-month forecast horizon - not a single yield-only model outperforms the random walk. Despite this, however, it proves to be very difficult to eliminate these models from the final Model Confidence Set. Only in rare occasions do models get discarded, indicating a substantial degree of model uncertainty. A final interesting observation to make from Tables 3 to 6 is that the two-step Nelson-Siegel models, regardless of whether these incorporate macroeconomic information or not, perform poorly across maturities and forecast horizons. This appears to contradict the results of Diebold and Li (2006) who find that the Nelson-Siegel model, especially the NS2-AR model, forecasts particularly well during the 1994-2000 period. As we will show below, the Nelson-Siegel model turns out to be one of the most prominent examples of the extent to which the forecast accuracy of term structure models can vary over time.

To further gauge the degree of model uncertainty, we analyze Cumulative Squared Prediction Error graphs. Because we construct forecasts for the entire sample period 1989 - 2003, we first take a step back and discuss results for the entire fifteen-year out-of-sample forecast period. The reason for doing this is that it also allows us to analyze our five-year training period. We feel this is interesting because it can give us some insights in the initial forecast combination weights, but more importantly, because the training period contains the 1990-1991 recession. Figures 4 to 7 show CSPEs for yield-only and macro models separately for each forecast horizon.¹³ Each line in the graph represents a different model and shows how that particular model performs relative to the random walk benchmark. In particular, an increasing CSPE indicates outperformance whereas a decreasing CSPE indicates that the random walk is making smaller forecasting errors.

As shown by the yellow bars in Figures 4 to 7, our out-of-sample period contains two NBER recessions. Both these recessions are characterized by a steep decline in short term interest rates as the Fed lowered its target interest rate, and by a sharp increase in the spread between long and short rates, see Figure 1(b). As it is also evident from earlier recessions, shown in Figure 1(a), spreads tend to remain high for quite a while until the Fed starts to raise short term interest rates again. The period in between the 1990-1991 and 2001 recessions, in particular the period 1994-2000, looks quite different on the other hand with much more

¹³To try and keep the number of graphs down we only show Trace CSPE graphs here. Graphs for individual maturities are available upon request.

stable interest rate dynamics, and seems best described as a low-volatility, low-spread regime for interest rates. Interestingly, Duffee (2002), Ang and Piazzesi (2003), and Diebold and Li (2006), among others, all tend to report a fair amount of predictability for this period. The CSPE graphs allow us to examine in much more detail how models perform during this period as well as during both recession periods, virtually on a month-to-month basis. Similar to us, Mönch (2008) and Carriero, Kapetanios, and Marcellino (2009) compare the forecast performance of a range of different models. They find that their preferred FAVAR and BVAR model, respectively, have the best relative RSME performance. To check the robustness of this result, they perform subsample analysis. However, both studies do so by considering just two subsamples, so we can still only judge models based on a single summary statistic for each subsample. This again does not give any real insight into where and why models perform well or not.

Although our out-of-sample period only contains two recessions, we believe the CSPE graphs reveal four important features. First, macro models perform better just prior to and during recessions. The CSPE lines are increasing in those periods, indicating that macro models forecast more accurately than the random walk. This is particularly true for long forecast horizons, see for example Figure 7. As several macro models simultaneously outperform the random walk, it clearly is the case that it is the macroeconomic information that is driving this result, and not so much any specific model. Ludvigson and Ng (2009) offer an interesting insight which can explain why macro information is useful in and around recessions. They find that macro factors explain risk premia much more than yield information does. Furthermore, they show that during recessions risk premia account for the largest portion of yield levels, implying that macro models will be better capable of forecasting the direction of yields in and around recessions. This certainly seems to be the case judging from Figures 4 to 7.

Second, most models perform poorly when the spread between long and short interest rates is high, after rates have begun to stabilize, but with medium-maturity yields being closer to short than they are to long rates. This is a typical shape of the term structure one or two years after recessions, in our case 1992-1993 and 2003. Only the AR-X models seems capable of coping this situation. Multivariate models all struggle in these periods. This is perhaps due to the fact that the larger number of estimated model parameters result leads to a less accurate fit of the term structure during these periods, which in turn is likely to lead to poor forecasts. Favero, Niu, and Sala (2009) offer some interesting insights on the role of estimation error on the forecasting performance of affine models, especially for longer-maturity yields. See also Duffee (2009) for comments on the numerical instability of affine models.

Third, yield-only models perform well in expansionary periods such as 1994-1998, corroborating the results in the above-mentioned studies, but very poorly in and around recession

periods.

Fourth, and this is our most important point, there is not a single model that clearly performs well across all maturities and forecast horizons. Hence there is a substantial degree of model uncertainty. Believing in any single model all the time can give very accurate forecasts in one period but, more troublesome, potentially very poor forecasts in other periods. Probably the best example of this is the Diebold and Li (2006) NS2-AR model. Figures 4 to 7 confirm the Diebold and Li (2006) results that the NS2-AR model gives very accurate forecasts for the period from 1994 to 2000, especially for longer forecast horizons. However, the CSPE graphs also show that most, if not all, of these forecast gains are confined to 1994 and 1995 when the NS2-AR model is by far the best performing model. During the years after 1995, the CSPE lines are all but flat, indicating that NS2-AR forecasts are about as accurate as the random walk model. Immediately following both the 1991 and 2001 recession, the NS2-AR performs by far the worst out of all models, as evidenced by the precipitous drop in CSPEs. All in all, the NS2-AR model is a prime example of the degree to which the forecast accuracy of term structure models can vary over time. Mönch (2008) also notes that "... some of the strong forecast performance of the Nelson-Siegel model documented by Diebold and Li may be due to their choice of forecast period."

Because in the end our main focus is on the 1994-2003 out-of-sample period, we show CSPEs in Figure 8 to 11 for the 1994-2003 period in the left-hand side and middle panels for individual models. These graphs therefore cover the same period as in Tables 3-6 and exclude the 1991 recession.¹⁴ In the next section we will confront these graphs with CSPE graphs based on forecast combinations, the right-hand side panels.

5 Forecast combination

Our cumulative squared prediction error analysis reveals that it seems virtually impossible to identify a single model that consistently outperforms the random walk for an entire out-of-sample period. The forecasting ability of individual models clearly varies over time considerably. Each model appears to play a complementary role in approximating the data generating process, at least during subperiods. Model uncertainty is troublesome if one has hopes of obtaining a single model for forecasting. A worthwhile endeavor for cushioning the effects of model uncertainty is to *combine* the forecasts of different models, see Timmermann (2006) for a recent survey. For example, one "solution" as to whether to impose no-arbitrage restrictions or not is to simply combine the forecasts from no-arbitrage models with those from unrestricted models. In this section we therefore examine several forecast combination schemes. Two combination methods are standard approaches which combine forecasts from

¹⁴Note that Figures 8 to 11 contain the same information for the 1994-2003 period as do Figures 4 to 7 do. However, the graphs differ because the CSPEs start at zero in 1989 and 1994, respectively.

all available models. In the third scheme we first filter out the worst-performing individual models before combining the forecasts from the remaining models. Below, we first discuss the different schemes. We then examine the forecast combination results and compare these with the single-model results for the 1994 - 2003 out-of-sample period.

5.1 Forecast combination schemes

Assuming we are combining forecasts from M different forecast models, a combined forecast for a h -month horizon for the yield with maturity τ_i is given by $\hat{y}_{T+h|T}^{(\tau_i)} = \sum_{m=1}^M w_{T+h|T,m}^{(\tau_i)} \hat{y}_{T+h|T,m}^{(\tau_i)}$, where $w_{T+h|T,m}^{(\tau_i)}$ denotes the weight assigned to the time- T forecast from the m^{th} model; $\hat{y}_{T+h|T,m}^{(\tau_i)}$.

Scheme 1: Equally weighted forecasts

The first forecast combination method we consider assigns equal weights to the forecasts from all individual models, i.e. $w_{T+h|T,m}^{(\tau_i)} = 1/M$ for $m = 1, \dots, M$. We denote the resulting combined forecast as Forecast Combination - Equally Weighted (**FC-EW**). As explained in Timmermann (2006), this approach is likely to work well if forecast errors from different models have similar variances and are highly correlated. Unreported statistics confirm that forecast errors from the individual models are indeed highly correlated here and have high variance.

Scheme 2: Inverted MSPE-weighted forecasts

The second forecast combination scheme we examine uses weights which are based on relative historical performance. More specifically, model weights are based on each model's (inverted) MSPE, relative to those of all other models, computed over a window of the previous v months. We denote these performance-based combinations forecasts by Forecast Combination - MSPE (**FC-MSPE**).¹⁵ The weight for model m is computed as $w_{T+h|T,m}^{(\tau_i)} = \frac{1/\text{MSPE}_{h|T,m}^{(\tau_i)}}{\sum_{m=1}^M (1/\text{MSPE}_{h|T,m}^{(\tau_i)})}$ where $\text{MSPE}_{h|T,m}^{(\tau_i)} = \frac{1}{v} \sum_{r=1}^v (\hat{y}_{T-r+1|T-h-r+1,m}^{(\tau_i)} - y_{T-r+1}^{(\tau_i)})^2$. A model with a lower MSPE is given a relatively larger weight than a worse performing model, see Timmermann (2006) for a discussion and Stock and Watson (2004) for an application to forecasting GDP growth. The weights applied in this and the previous forecast combination scheme are always bounded between 0 and 1. Other approaches for which this does not necessarily need to be the case, in particular OLS-based weights (see again Timmermann, 2006), proved to be problematic here due to multicollinearity problems among the different forecasts. This

¹⁵Note that whereas in Panels A and B of Tables 3 to 6 we report results for the *Root* MSPE, Timmermann (2006) argues that it is better to use the MSPE to construct model weights. We therefore use MSPE in this forecast combination scheme.

resulted in often extreme (offsetting) weights and we therefore decided not to further pursue these approaches.

The number that should be used for v is difficult to determine *a priori*. Using a smaller window will make weights more responsive to changes in models' forecasting accuracy, but at the same time it will also tend to make them more noisy. The optimal choice of v will therefore need to be determined empirically. Somewhat counterintuitive maybe, we found that using an expanding window approach works the best. We tried different lengths in a moving window approach (in particular, $v = 12, 24$ and 60 months) but for shorter windows results were (marginally) worse. Similarly, a weighted approach using declining weights for older forecast errors as in Diebold and Pauly (1987) also gave worse results. We settled on using an expanding window, whose length is initially set to $v = 60$ months but which increases with every new yield realization that becomes available.

Finally, for Scheme 1 and 2 we distinguish between using forecasts of macro models only; **FC-EW-X** and **FC-MSPE-X**, and combining forecasts across all models; **FC-EW-ALL** and **FC-MSPE-ALL**.

Scheme 3: Trimming via Model Confidence Set

The Model Confidence Set approach for evaluating forecast performance, as described in Section 4, can also be implemented as an initial model elimination mechanism. The idea of trimming the available set of models prior to combining forecasts has been proposed in several studies, see for example Timmermann (2006). As these studies show, trimming is an efficient way to first discard of the “worst” models, and then combine the forecasts from the surviving models. The MCS approach seems particularly suitable for doing so because it requires few *a priori* decisions, such as for example having to select the number of remaining models. As the MCS grows and shrinks over time, so does the number of models whose forecasts are combined into a single number.

As our third and final forecast combination scheme we therefore combine forecasts from models that survive the MCS approach, using equal and MSPE-based weights. Specifically, in order to construct a h -month ahead combination forecast at time T , we use the T_{SQ} statistic to construct $\widehat{\mathcal{M}}_{0,25}^*$, using a confidence level of $1 - \alpha = 75\%$.¹⁶ We determine $\widehat{\mathcal{M}}_{0,25}^*$ using an expanding window of previous forecasts, starting from the initial sixty-month sample 1989:1 - 1993:12. To determine the MCS we always start by inserting all available individual models, i.e. the entire set M_0 , so as not to have to make a (subjective) initial model selection.

By studying which models are contained in $\widehat{\mathcal{M}}_{0,25}^*$ over time we can also again infer information about the consistency of models' forecasting performance. In Tables 3 to 6

¹⁶We also implement this combination scheme with the Range and Deviation statistics as in Hansen, Lunde, and Nason (2003), as well as for a 90% confidence level. Results were very similar to marginally worse than the statistics we report in Panels C of Tables 3 to 6.

we therefore also report the percentage of times each individual model is included in the model confidence sets for the 1994:1 - 2003:12 sample (in parentheses below each (relative) (T)RMSPE statistic). For example, a number close to one indicates that a model is nearly always included in the combination set, whereas a number close or equal to zero shows that it is typically excluded.¹⁷

5.2 Forecast combination results

We feel that there are at least three important conclusions that we can draw from the forecast combination results in Tables 3-6 and Figures 8 to 11. First, several forecast combination schemes perform better or similar to the random walk across different forecast horizons. TRMSPEs and RMSPEs are often below one, albeit marginally in some cases. Compared to the best performing individual models, prediction errors for the forecast combination schemes are somewhat higher, but they certainly seem to be more stable, as is evident from the right-hand side panels in Figures 8 to 11, even though it may not be initially clear from Figures 12 to 15. Focusing on Figures 8 to 11, the performance variability associated with macro models is reduced substantially in the first part of the sample and the bad performance of yield-only models during and after the 2001 recession is mitigated.

Second, averaging across all the models, after trimming out the worst performing ones using the MCS approach, gives the best performance for shorter forecast horizons (the bottom two lines in the tables). The gains are particularly encouraging for shorter maturities. The inclusion percentages (in parentheses in the tables) reveal that this trimming-via-MCS scheme nearly always select the best performing individual models in the forecast combination. For example, with a 1-month horizon for the 1-month maturity, the VAR, VAR-X, ATSM, ATSM-X, and the NS2-VAR-X models, are basically always included. In other cases, such as for example for the same horizon but now for the 1-year maturity, only a single model (AR-X) actually makes it into the MCS. The differences between using equal and MSPE weights are minor, enforcing the conclusion that it is the trimming procedure which is most beneficial to the forecast combination method, not so much which weights are used to sum the individual forecasts. The light-blue lines in Figures 8 and 9 show that the FC-MCS-EW scheme does not always necessarily provide the best forecasts, but it certainly produces much less prediction error volatility.

Third, averaging only across macro models produces the most accurate forecasts for longer horizons. The MCS inclusion percentages indicate that it is very difficult to discard

¹⁷Note that the MCS to compute the inclusion percentages in Tables 3 to 6 is based on an expanding window that starts in 1989:1, whereas the full-sample MCS results in those same tables are based on the sample 1994:1 - 2003:12. It can happen therefore that a model is included in the full-sample MCS, while at the same time it is hardly ever included in the expanding MCS trimming combination scheme, i.e. it has a percentage close to, or equal to, zero.

models and many specifications indeed have high inclusion percentages. However, combining forecasts of macro models only, lines 4 in Panel C in each table, gives RMSPE ratios which are almost always below one, in particular for the twelve-month horizon. The FC-MSPE-X scheme appears to be the best forecasting strategy. Figures 10 and 11 reveal that that is in part due to the fact that the FC-MSPE-X performs best during and after the 2001 recession. Nevertheless, the results suggest that the past performance of individual models provides a useful insight as to which models to include in the forecast combination. Going back to Table 6, the FC-MSPE-X scheme does particularly well across maturities for the twelve-month horizon but, quite important, especially for longer-dated maturities. Earlier studies with individual models tend to find that it is typically very hard to accurately forecast long-term rates and our results in Panel A and B confirm this. The outperformance of the FC-MSPE-X scheme relative to the random walk is 8% for the ten-year maturity whereas the best individual model (AR-X) is only barely below one. This result suggest that forecasting combinations can potentially be very useful for forecasting long-maturity yields with long forecast horizons.

6 Conclusion

This paper addresses the task of forecasting the term structure of interest rates. Several recent studies have shown that significant steps forward are being made in this area. We contribute to the existing literature by further assessing the importance of incorporating macroeconomic information, and, in particular, by examining model uncertainty. Our results show that incorporating macroeconomic information indeed helps to improve forecasts of individual models. Our main result, however, is that the predictive performance of individual models can be strongly time-varying, which makes putting all one's eggs in a single model basket risky. Our suggested alternative, combining forecasts across different models, not only mitigates model uncertainty, but also results in accurate forecasts.

We have examined the forecast accuracy of a range of models with varying degrees of complexity. We showed that the predictive ability of individual models varies over time considerably. Models that incorporate macroeconomic variables are more accurate during interest rate regimes where the uncertainty about the future path of interest rates is substantial. As an example we mention the period during and after the 2001 recession. Models without macro information do particularly well in subperiods where the term structure has a more stable pattern (such as in the late 1990s) or when the spread between long and short-maturity yield closes.

The fact that different models forecast well in different subperiods confirms ex-post that alternative model specifications play a complementary role in approximating the data generating process. We believe our results provide a strong claim for using forecast combination

techniques as an alternative to believing in a single model. We show that combining forecasts of all individual models with and without macro factors, after trimming out the worst performing models using the Model Confidence Set approach, gives accurate forecasts for short forecast horizons. Combination forecasting of models with macro information, using a weighting method that is based on relative historical performance over a long sample, results in superior forecasts for long forecast horizons. The gains in the latter case are particularly encouraging for longer-dated maturities, which have proven to be notoriously difficult to predict.

References

- Ang, A. and G. Bekaert (2002), Regime Switches in Interest Rates, *Journal of Business & Economic Statistics*, **20**, 163–182.
- Ang, A., S. Dong, and M. Piazzesi (2007), No-Arbitrage Taylor Rules, *NBER Working paper 13448*.
- Ang, A. and M. Piazzesi (2003), A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, *Journal of Monetary Economics*, **50**, 745–787.
- Ang, A., M. Piazzesi, and M. Wei (2006), What does the Yield Curve tell us about GDP Growth?, *Journal of Econometrics*, **131**, 359–403.
- Backus, D., S. Foresi, and C. Telmer (1998), Discrete-Time Models of Bond Pricing, *NBER Working paper 6736*.
- Bansal, R. and H. Zhou (2002), Term Structure of Interest Rates with Regime Shifts, *Journal of Finance*, **57**, 1997–2043.
- Bernanke, B. S., J. Boivin, and P. Elias (2005), Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics*, **120**, 387–422.
- Bikbov, R. and M. Chernov (2008), No-Arbitrage Macroeconomic Determinants of the Yield Curve, *Working paper, London Business School*.
- Bliss, R. R. (1997), Testing Term Structure Estimation Methods, *Advances in Futures and Options Research*, **9**, 197–231.
- Carriero, A., G. Kapetanios, and M. Marcellino (2009), Forecasting Government Bond Yields, *Working paper University of London*.
- Chen, R. R. and L. Scott (1993), Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates, *Journal of Fixed Income*, **3**, 14–31.
- Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch (2009), The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models, *Working paper, University of Pennsylvania and Federal Reserve Bank of San Francisco*.
- Christoffersen, P. F. and F. X. Diebold (1998), Cointegration and Long-Horizon Forecasting, *Journal of Business & Economic Statistics*, **16**, 450–458.

- Cox, J., J. E. Ingersoll, and S. A. Ross (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*, **53**, 385–407.
- Croushore, D. (2006), Forecasting with Real-Time Macroeconomic Data, in G. Elliot, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland, 961–982.
- Dai, Q. and T. Philippon (2006), Fiscal Policy and the Term Structure of Interest Rates, *NBER Working paper*.
- Dai, Q. and K. J. Singleton (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance*, **55**, 1943–1978.
- Dai, Q., K. J. Singleton, and W. Yang (2007), Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields, *Review of Financial Studies*, **20**, 1669–1706.
- DeWachter, H. and M. Lyrio (2006), Macro Factors and the Term Structure of Interest Rates, *Journal of Money, Credit and Banking*, **38**, 119–140.
- Diebold, F. X. and C. Li (2006), Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, **130**, 337–364.
- Diebold, F. X. and P. Pauly (1987), Structural Change and the Combination of Forecasts, *Journal of Forecasting*, **6**, 21–40.
- Diebold, F. X., G. D. Rudebusch, and B. Aruoba (2006), The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach, *Journal of Econometrics*, **131**, 309–338.
- Duffee, G. R. (2002), Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance*, **5**, 405–443.
- Duffee, G. R. (2009), Forecasting with the Term Structure: The Role of No-arbitrage Restrictions, *Working paper, Johns Hopkins University*.
- Duffie, D. and R. Kan (1996), A Yield-Factor Model of Interest Rates, *Mathematical Finance*, **6**, 379–406.
- Evans, C. L. and D. Marshall (1998), Monetary Policy and the Term Structure of Nominal Interest Rates: Evidence and Theory, *Carnegie-Rochester Conference Series on Public Policy*, **49**, 53–111.
- Evans, C. L. and D. Marshall (2007), Economic Determinants of the Nominal Treasury Yield Curve, *Journal of Monetary Economics*, **54**, 1986–2003.
- Fama, E. F. and R. R. Bliss (1987), The Information in Long-Maturity Forward Rates, *American Economic Review*, **77**, 680–692.
- Favero, C. A., L. Niu, and L. Sala (2009), Term Structure Forecasting: No-arbitrage Restrictions versus Large Information Set, *Working paper, Bocconi University*.
- Gürkanyak, R. S., B. Sack, and J. H. Wright (2007), The U. S. Treasury Yield Curve: 1961 to the Present, *Journal of Monetary Economics*, **54**, 2291–2304.
- Hansen, P. R., A. Lunde, and J. M. Nason (2003), Choosing the Best Volatility Model: The Model Confidence Set Approach, *Oxford Bulletin of Economics and Statistics*, **65**, 839–861.

- Hansen, P. R., A. Lunde, and J. M. Nason (2005), Model Confidence Sets for Forecasting Models, *Federal Reserve Bank of Atlanta, Working paper 2005-7*.
- Hordahl, P., O. Tristani, and D. Vestin (2006), A Joint Econometric Model of Macroeconomic and Term-Structure Dynamics, *Journal of Econometrics*, **131**, 405–444.
- Kim, D. H. and J. H. Wright (2005), An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates, *Working paper, Federal Reserve Board Washington*, **33**.
- Litterman, R. and J. Scheinkman (1991), Common Factors Affecting Bond Returns, *Journal of Fixed Income*, 54–61.
- Ludvigson, S. C. and S. Ng (2009), Macro Factors in Bond Risk Premia, *forthcoming in Review of Financial Studies*.
- Marcellino, M., J. H. Stock, and M. W. Watson (2006), A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series, *Journal of Econometrics*, **127**, 499–526.
- Mönch, E. (2006), Term Structure Surprises: The Predictive Content of Curvature, Level, and Slope.pdf, *Working paper, Humboldt University Berlin*.
- Mönch, E. (2008), Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach, *Journal of Econometrics*, **146**, 24–43.
- Nelson, C. R. and A. F. Siegel (1987), Parsimonious Modeling Of Yield Curves, *Journal of Business*, **60**, 473–489.
- Piazzesi, M. (2003), Affine Term Structure Models, in Y. Ait-Sahalia and L. Hansen (eds.), *forthcoming in Handbook of Financial Econometrics*, Elsevier.
- Politis, D. N. and J. P. Romano (1994), The Stationary Bootstrap, *Journal of the American Statistical Association*, **89**, 1303–1313.
- Rudebusch, G. D. and T. Wu (2008), A Macro-Finance Model of the Term-Structure, Monetary Policy, and the Economy, *Economic Journal*, **118**, 1–21.
- Stock, J. H. and M. W. Watson (2002a), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics*, **20**, 147–162.
- Stock, J. H. and M. W. Watson (2002b), Forecasting Using Principal Components From a Large Number of Predictors, *Journal of the American Statistical Association*, **97**, 1167–1179.
- Stock, J. H. and M. W. Watson (2004), Combining Forecasts of Output Growth in a Seven-Country Data Set, *Journal of Forecasting*, **23**, 405–430.
- Stock, J. H. and M. W. Watson (2005), Implications of Dynamic Factor Models for VAR Analysis, *Working paper*.
- Svensson, L. E. O. (1994), Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994, *NBER Working Paper Series*, **4871**.
- Timmermann, A. (2006), Forecast Combinations, in G. Elliot, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland, 135-196.

Vasicek, O. A. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, **5**, 177–188.

Welch, I. and A. Goyal (2008), A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies*, **21**, 1455–1508.

Wu, T. (2005), Macro Factors and the Affine Term Structure of Interest Rates, *Working paper*, *Federal Reserve Bank of San Francisco*.

A Individual Interest Rate Models

In this appendix we provide some further details on how we perform inference on the parameters of each of the models in Section 3.

A.1 AR model

We estimate the parameters $\{c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)}\}$ using standard ordinary least squares (OLS). Given the parameter estimates, we construct iterated forecasts as

$$\hat{y}_{T+h}^{(\tau_i)} = \hat{c}^{(\tau_i)} + \hat{\phi}^{(\tau_i)} \hat{y}_{T+h-1}^{(\tau_i)} + \hat{\psi}^{(\tau_i)'} \hat{X}_{T+h} \quad (\text{A-1})$$

with $\hat{y}_T^{(\tau_i)} = y_T^{(\tau_i)}$. We construct forecasts from the AR model both with and without macroeconomic factors. The macro factor forecasts, \hat{X}_{T+h} , are iterated forecasts constructed from the VAR(3) macro factor model.

A.2 VAR model

We estimate the equation parameters $\{c, \Phi, \Psi\}$ in (5) using equation-by-equation OLS as each equation has an identical set of regressors. We construct forecasts as follows:

$$\hat{Y}_{T+h} = \hat{c} + \hat{\Phi} \hat{F}_{T+h-1} + \hat{\Psi} \hat{X}_{T+h} \quad (\text{A-2})$$

where we compute the yield factor forecasts, \hat{F}_{T+h-1} , by first calculating the principal component factor loadings using data only up until month T and then multiplying these loadings with the iterated yields forecasts.

A.3 Nelson-Siegel model

We estimate the Nelson-Siegel model with the two-step approach of Diebold and Li (2006) as well as the one-step approach of Diebold, Rudebusch, and Aruoba (2006).

In the two-step approach we fix λ to 16.42, which, as shown in Diebold and Li (2006), maximizes the curvature factor loading at a 30-month maturity. Given the value for λ we then estimate the vector of latent factors for every individual month by applying OLS to the cross-section of yields (all 18 maturities). From this first step we obtain time-series for the three factors, $\{\beta_t\}_{t=1}^T$. The second step consists of estimating the dynamics of the factors in (7) by either fitting a single VAR(1) model, or by separate AR(1) models.

In the one-step approach we estimate the unknown parameters and latent factors by means of the Kalman Filter. We maximize the likelihood using the prediction error decomposition of the state space model in (6) and (7). For each sample in the recursive estimation procedure, we first run the two-step approach with a VAR(1) specification for the state vector to obtain starting values. The unconditional mean and covariance matrix of $\{\beta_t\}_{t=1}^T$ are used to start the Kalman Filter. We discard the first 12 observations when evaluating the likelihood. All variance parameters of the diagonal matrix H and the full matrix Q are initialized to 1. The covariance terms in Q are initialized to 0. In the optimization procedure, we maximize the likelihood by treating the standard deviations as parameters instead of optimizing over the variance parameters directly, to ensure that all variance parameters are positive. We initialize λ to 16.42.

We obtain iterated forecasts for the factors as follows:

$$\hat{f}_{T+h} = \hat{a} + \hat{\Gamma} \hat{f}_{T+h-1} \quad (\text{A-3})$$

where $\hat{f}_{T+h} = (\hat{\beta}_{1,T+h}, \hat{\beta}_{2,T+h}, \hat{\beta}_{3,T+h})'$ for the model without macro factors, whereas $\hat{f}_{T+h} = (\hat{\beta}_{1,T+h}, \hat{\beta}_{2,T+h}, \hat{\beta}_{3,T+h}, \hat{M}_{T+h}, \hat{M}_{T+h-1}, \hat{M}_{T+h-2})'$ when macro factors are included. The factor forecasts are then inserted in the measurement equation to compute interest rate forecasts:

$$\hat{y}_{T+h}^{(\tau_i)} = \hat{\beta}_{1,T+h} + \hat{\beta}_{2,T+h} \left(\frac{1 - \exp(-\tau_i/\hat{\lambda})}{\tau_i/\hat{\lambda}} \right) + \hat{\beta}_{3,T+h} \left(\frac{1 - \exp(-\tau_i/\hat{\lambda})}{\tau_i/\hat{\lambda}} - \exp(-\tau_i/\hat{\lambda}) \right) \quad (\text{A-4})$$

A.4 Affine model

To estimate the affine model we assume that every yield is contaminated with measurement error. We estimate the parameters in the resulting state space model by applying the two-step approach used in Ang, Piazzesi, and Wei (2006). We make the latent factors Z_t observable by extracting the first three principal components from the panel of yields. The first step of the estimation procedure consists of estimating the equation and variance parameters of the state equations in (23). In the second step we estimate the remaining parameters $\{\delta_0, \delta_1, \lambda_0, \lambda_1\}$. We first estimate $\{\delta_0, \delta_1\}$ by applying OLS to the short rate equation (13) where we use the 1-month yield as the observable short rate. We then estimate the risk premia parameters $\{\lambda_0, \lambda_1\}$ by minimizing the sum of squared yields errors, taking as given the parameter estimates from the first step, $\{\hat{\mu}, \hat{\Psi}, \hat{\Sigma}\}$ and the short rate parameters $\{\hat{\delta}_0, \hat{\delta}_1\}$. When we optimize over the risk premium parameters in the second step, we initialize all risk premia parameters with zeros. Common approaches for obtaining starting values for the risk premia parameters which tend to first estimate either λ_0 or λ_1 in a separate step, gave unsatisfactory results. So we decided to initialize the optimization procedure assuming that *all* risk premium parameters are zero. We incorporate macro factors by writing the state equations in companion form. All parameters in the short rate equations and the time-varying risk premia that are associated with lags of the macro factors are set to zero.

Yield forecasts are generated by forward iteration of the state equations:

$$\hat{f}_{T+h} = \hat{\mu} + \hat{\Psi} \hat{f}_{T+h-1} \quad (\text{A-5})$$

where $\hat{f}_{T+h} = \hat{Z}_{T+h}$ for the yields-only model whereas $\hat{f}_{T+h} = (\hat{Z}_{T+h}, \hat{M}_{T+h-1}, \hat{M}_{T+h-2})$ for the affine model with macro factors.

With the estimated parameters substituted in the recursive bond pricing coefficient equations $a^{(\tau_i)}$ and $b^{(\tau_i)}$, we then construct interest rate forecasts as

$$\hat{y}_{T+h}^{(\tau_i)} = \hat{a}^{(\tau_i)} + \hat{b}^{(\tau_i)} \hat{f}_{T+h} \quad (\text{A-6})$$

B Forecast Evaluation Criteria

In the tables below we report the (Trace) Root Mean Squared Prediction Errors. Given a sample of R out-of-sample forecasts with a h -month ahead forecast horizon, we compute the RMSPE for a τ_i -month yield for model m , with $m = 1, \dots, M$, as follows:

$$\text{RMSPE}_m(\tau_i) = \sqrt{\frac{1}{R} \sum_{t=1}^R \left(\hat{y}_{t+h|t,m}^{(\tau_i)} - y_{t+h}^{(\tau_i)} \right)^2} \quad (\text{B-1})$$

where $y_{t+h}^{(\tau_i)}$ is the yield for a τ_i -month maturity observed at time $t+h$, while $\hat{y}_{t+h|t,m}^{(\tau_i)}$ is its model- m forecast, made at time t .

The TRMSPE is an aggregate over all N yield maturities:

$$\text{TRMSPE}_m = \sqrt{\frac{1}{N} \frac{1}{R} \sum_{i=1}^N \sum_{t=1}^R \left(\hat{y}_{t+h|t}^{(\tau_i)} - y_{t+h}^{(\tau_i)} \right)^2} \quad (\text{B-2})$$

The Cumulative Squared Prediction Error (CSPE) computes the sum of squared prediction errors for a model m , relative to those of a benchmark model, here the random walk (RW):

$$\text{CSPE}_{m,T}(\tau_i) = \sum_{t=1}^T \left[\left(\hat{y}_{t+h|t,\text{RW}}^{(\tau_i)} - y_{t+h}^{(\tau_i)} \right)^2 - \left(\hat{y}_{t+h|t,m}^{(\tau_i)} - y_{t+h}^{(\tau_i)} \right)^2 \right] \quad (\text{B-3})$$

If a model outperforms the random walk, then $\text{CSPE}_{m,T}$ will be an increasing series. If the random walk produces more accurate forecasts, then $\text{CSPE}_{m,T}$ will tend to be decreasing. The CSPE is informative at each point in time basically, as it will go up in any given month if the model outperforms its benchmark, whereas it will go down vice versa.

Table 1: Summary statistics

maturity	mean	stdev	skew	kurt	min	max	JB	ρ_1	ρ_{12}	ρ_{24}
1-month	6.049	2.797	0.913	4.336	0.794	16.162	85.671	0.968	0.690	0.402
3-month	6.334	2.896	0.871	4.237	0.876	16.020	76.380	0.974	0.708	0.415
6-month	6.543	2.927	0.788	4.016	0.958	16.481	58.796	0.976	0.723	0.444
1-year	6.755	2.860	0.661	3.763	1.040	15.822	38.907	0.975	0.733	0.474
2-year	7.032	2.724	0.644	3.672	1.299	15.650	35.240	0.978	0.748	0.526
3-year	7.233	2.594	0.685	3.663	1.618	15.765	38.796	0.979	0.763	0.560
4-year	7.392	2.510	0.728	3.607	1.999	15.821	41.640	0.980	0.771	0.582
5-year	7.483	2.449	0.759	3.478	2.351	15.005	42.454	0.982	0.786	0.607
6-year	7.611	2.406	0.791	3.437	2.663	14.979	45.236	0.983	0.797	0.626
7-year	7.659	2.344	0.841	3.488	3.003	14.975	51.562	0.983	0.787	0.623
8-year	7.728	2.320	0.841	3.365	3.221	14.936	49.798	0.984	0.809	0.651
9-year	7.767	2.317	0.877	3.427	3.389	15.018	54.765	0.985	0.813	0.656
10-year	7.745	2.266	0.888	3.496	3.483	14.925	57.117	0.985	0.796	0.647

Notes: The table shows summary statistics for our sample of end-of-month continuously compounded U.S. zero-coupon yields. Reported are the mean, standard deviation, skewness, kurtosis, minimum, maximum, the Jarque-Bera test statistic for normality and the 1st, 12th and 24th sample autocorrelation. The results shown are for annualized yields (in percentage points). The sample period is January 1970 - December 2003 (408 monthly observations).

Table 2: Macroeconomic dataset

group (code)	transformation code	description
real output and income		
1	7	Personal Income (B\$, Chain 2000)
1	7	Personal Income Less Transfer Payments (B\$, Chain 2000)
1	7	Industrial Production Index - Total Index
1	7	Industrial Production Index - Products, Total
1	7	Industrial Production Index - Final Products
1	7	Industrial Production Index - Consumer Goods
1	7	Industrial Production Index - Durable Consumer Goods
1	7	Industrial Production Index - Nondurable Consumer Goods
1	7	Industrial Production Index - Business Equipment
1	7	Industrial Production Index - Materials
1	7	Industrial Production Index - Durable Goods Materials
1	7	Industrial Production Index - Nondurable Goods Materials
1	7	Industrial Production Index - Manufacturing
1	7	Industrial Production Index - Residential Utilities
1	7	Industrial Production Index - Fuels
1	1	NAPM Production Index (percent)
1	8	Manufacturing Capacity Utilization
employment and hours		
2	1	Index of Help-Wanted Advertising In Newspapers (1967=100, SA)
2	1	Employment Ratio of Help-Wanted Ads to No. of Unemployed in Civilian Labor Force
2	7	Civilian Labor Force: Employed, Total (Thousands, SA)
2	7	Civilian Labor Force: Employed, Nonagricultural Industries (Thousands, SA)
2	1	Unemployment Rate: All Workers, 16 Years & Over (percent, Sa)
2	8	Unemployment by Duration: Average (Mean) Duration In Weeks (SA)
2	7	Unemployment by Duration: Persons Unempl. Less Than 5 Weeks (Thousands, SA)
2	7	Unemployment by Duration: Persons Unemployment 5 To 14 Weeks (Thousands, SA)
2	7	Unemployment by Duration: Persons Unemployment 15 Weeks or more (Thousands, SA)
2	7	Unemployment by Duration: Persons Unemployment 15 To 26 Weeks (Thousands, SA)
2	7	Unemployment by Duration: Persons Unemployment 27 Weeks or more (Thousands, SA)
2	7	Average Weekly Initial Claims, Unemployment Insurance (Thousands)
2	7	Employees on Nonfarm Payrolls: Total Private
2	7	Employees on Nonfarm Payrolls - Goods-Producing
2	7	Employees on Nonfarm Payrolls - Mining
2	7	Employees on Nonfarm Payrolls - Construction
2	7	Employees on Nonfarm Payrolls - Manufacturing
2	7	Employees on Nonfarm Payrolls - Durable Goods
2	7	Employees on Nonfarm Payrolls - Nondurable Goods
2	7	Employees on Nonfarm Payrolls - Service-Providing
2	7	Employees on Nonfarm Payrolls - Trade, Transportation, And Utilities
2	7	Employees on Nonfarm Payrolls - Wholesale Trade
2	7	Employees on Nonfarm Payrolls - Retail Trade
2	7	Employees on Nonfarm Payrolls - Financial Activities
2	7	Employees on Nonfarm Payrolls - Government
2	7	Employee Hours in Nonagricultural Establishments (B. Hours)
2	1	Avg Wkly Hrs of Prod or Nonsup Workers on Priv. Nonfarm Payrolls: Goods-Producing
2	8	Avg Wkly Hrs of Prod or Nonsup Workers on Priv. Nonfarm Payrolls: Manufacturing Overtime Hours
2	1	Average Weekly Hours, Manufacturing (hours)
2	1	NAPM Employment Index (percent)
real retail		
3	7	Sales of Retail Stores (M\$, Chain 2000)
manufacturing and trade sales		
4	7	Manufacturing and Trade Sales (M\$, Chain 1996)
consumption		
5	7	Real Consumption: a0m224/gmdc (a0m224 is from TCB)
housing starts and sales		
6	4	Housing Starts: Nonfarm (1947-58); Total Farm & Nonfarm (1959-) (Thousands of Units, SAAR)
6	4	Housing Starts: Northeast (Thousands of Units, SA)
6	4	Housing Starts: Midwest (Thousands of Units, SA)
6	4	Housing Starts: South (Thousands of Units, SA)
6	4	Housing Starts: West (Thousands of Units, SA)
6	4	Housing Authorized: Total New Private Housing Units (Thousands of Units, SAAR)
6	4	Houses Authorized by Building Permits: Northeast (Thousands of Units, SA)
6	4	Houses Authorized by Building Permits: Midwest (Thousands of Units, SA)
6	4	Houses Authorized by Building Permits: South (Thousands of Units, SA)
6	4	Houses Authorized by Building Permits: West (Thousands of Units, SA)

Table 2: Macroeconomic dataset (continued)

group (code)	transformation code	description
inventories		
7	1	NAPM Inventories Index (percent)
7	7	Manufacturing and Trade Inventories (B\$, Chain 2000)
7	8	Ratio of Manufacturing and Trade Inventories to Sales (\$, Chain 2000)
orders		
8	1	Purchasing Managers' Index (SA)
8	1	NAPM New Orders Index (percent)
8	1	NAPM Vendor Deliveries Index (percent)
8	7	Manufacturers' New Orders, Consumer Goods And Materials (B\$, Chain 1982)
8	7	Manufacturers' New Orders, Durable Goods Industries (B\$, Chain 2000)
8	7	Manufacturers' New Orders, Nondefense Capital Goods (M\$, Chain 1982)
8	7	Manufacturers' Unfilled Orders, Durable Goods Industries (B\$, Chain 2000)
equities		
9	7	S&P's Common Stock Price Index: Composite (1941-43=10)
9	7	S&P's Common Stock Price Index: Industrials (1941-43=10)
9	8	S&P's Composite Common Stock: Dividend Yield (percent p.a.)
9	7	S&P's Composite Common Stock: Price-Earnings Ratio (percent, NSA)
exchange rates		
10	7	United States: Effective Exchange Rate (MERM model)(index number)
10	7	Foreign Exchange Rate: Switzerland (Swiss Franc per US\$)
10	7	Foreign Exchange Rate: Japan (Yen per US\$)
10	7	Foreign Exchange Rate: United Kingdom (US\$ per Sterling)
10	7	Foreign Exchange Rate: Canada (Canadian Dollar per US\$)
interest rates		
11	1	Interest Rate: Effective Federal Funds (percent p.a., NSA)
money and credit quantity aggregates		
12	7	Money Stock: M1 (B\$, SA)
12	7	Money Stock: M2 (B\$, SA)
12	7	Money Stock: M3 (B\$, SA)
12	7	Money Supply - M2 In 1996 Dollars
12	7	Monetary Base, Adj For Reserve Requirement Changes (M\$, SA)
12	7	Depository Inst Reserves:Total, Adj For Reserve Req Chgs (M\$, SA)
12	7	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs (M\$, SA)
12	7	Commercial & Industrial Loans Outstanding In 1996 Dollars
12	1	Weekly Report of Commercial Bank Lending: Net Change Commercial & Industrial Loans (B\$, SAAR)
12	7	Consumer Credit Outstanding - Nonrevolving
12	8	Ratio of Consumer Installment Credit To Personal Income (percent)
price indexes		
13	7	PPI: Finished Goods (1982=100, SA)
13	7	PPI: Finished Consumer Goods (1982=100, SA)
13	7	PPI: Intermediate Materials, Supplies & Components (1982=100, SA)
13	7	PPI: Crude Materials (1982=100, SA)
13	7	Spot market price index: BLS & CRB: all commodities (1967=100)
13	7	Index of Sensitive Materials Prices (1990=100)
13	1	NAPM Commodity Prices Index (percent)
13	7	CPI-U: All Items (1982-84=100, SA)
13	7	CPI-U: Apparel & Upkeep (1982-84=100, SA)
13	7	CPI-U: Transportation (1982-84=100, SA)
13	7	CPI-U: Medical Care (1982-84=100, SA)
13	7	CPI-U: Commodities (1982-84=100, SA)
13	7	CPI-U: Durables (1982-84=100, SA)
13	7	CPI-U: Services (1982-84=100, SA)
13	7	CPI-U: All Items Less Food (1982-84=100, SA)
13	7	CPI-U: All Items Less Shelter (1982-84=100, SA)
13	7	CPI-U: All Items Less Midical Care (1982-84=100, SA)
13	7	PCE, Implicit Price Deflator: PCE (1987=100)
13	7	PCE, Implicit Price Deflator: PCE; Durables (1987=100)
13	7	PCE, Implicit Price Deflator: PCE; Nondurables (1996=100)
13	7	PCE, Implicit Price Deflator: PCE; Services (1987=100)
average hourly earnings		
14	7	Avg Hourly Earnings of Prod or Nonsup Workers On Priv. Nonfarm Payrolls - Goods-Producing
14	7	Avg Hourly Earnings of Prod or Nonsup Workers On Priv. Nonfarm Payrolls - Construction
14	7	Avg Hourly Earnings of Prod or Nonsup Workers On Priv. Nonfarm Payrolls - Manufacturing
miscellaneous		
15	8	University of Michigan Index of Consumer Expectations

Notes: The table lists the individual macro series that we use to construct macro factors. The series are categorized in 15 groups: (1) real output and income, (2) employment and hours, (3) real retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) prices indices, (14) average hourly earnings and (15) miscellaneous. The transformations applied to original series are coded as: 1 \equiv no transformation (levels are used), 4 \equiv logarithm of the level, 7 \equiv annual first differences of the log levels and 8 \equiv annual first differences of the levels. The sample period is January 1970 - December 2003 (408 observations). Series are from the Global Insights Basic Economics Database^{3,2} and The Conference Board's Indicators Database. "[N]SA" stands from (Non-)Seasonally Adjusted whereas "SAAR" stands for Seasonally Adjusted Annual Rate.

Table 3: [T]RMSPE 1994:1 - 2003:12, 1-month forecast horizon

Models	[T]RMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	101.59	30.12	21.18**	21.82	25.71**	29.12*	30.48**	29.30**	27.95**
		[0.00]	[0.93]	[0.11]	[0.84]	[0.93]	[1.00]	[1.00]	[1.00]
Panel A: Models without macro factors									
AR	1.02	1.04	1.07*	1.06	1.05*	1.03	1.01**	1.01**	1.01**
		[0.00]	[0.77]	[0.00]	[0.30]	[0.43]	[0.69]	[0.79]	[0.48]
VAR	1.06	0.83**	1.03*	1.23	1.14	1.13	1.04**	1.05	1.11
		[1.00]	[0.19]	[0.00]	[0.00]	[0.00]	[0.04]	[0.03]	[0.25]
NS2-AR	1.10	0.94	1.13*	1.27	1.24	1.19	1.11*	1.06	1.07**
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.03]
NS2-VAR	1.04	0.94	0.96**	1.10	1.10	1.11	1.06**	1.03*	1.06*
		[0.00]	[0.90]	[0.00]	[0.00]	[0.00]	[0.00]	[0.48]	[0.02]
NS1	1.06	1.16	1.09	1.08	1.05**	1.10	1.07**	1.04	1.06*
		[0.00]	[0.07]	[0.00]	[0.13]	[0.00]	[0.15]	[0.80]	[0.35]
ATSM	1.07	0.84**	0.93**	1.15	1.23	1.18	1.04**	1.08	1.07*
		[0.99]	[0.84]	[0.00]	[0.00]	[0.00]	[0.69]	[0.00]	[0.38]
Panel B: Models with macro factors									
AR-X	0.99	0.98	0.95**	0.96**	0.98**	0.98**	0.99**	1.00**	0.99**
		[0.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
VAR-X	1.02	0.83**	0.99**	1.03	1.01**	1.12	1.02**	1.02**	1.03**
		[1.00]	[0.85]	[0.00]	[0.68]	[0.00]	[0.56]	[0.79]	[0.08]
NS2-AR-X	1.09	0.90	1.22	1.31	1.28	1.17	1.05**	1.06	1.06**
		[0.08]	[0.00]	[0.00]	[0.00]	[0.00]	[0.46]	[0.34]	[0.03]
NS2-VAR-X	1.05	0.83**	1.05**	1.17	1.20	1.13	1.03**	1.05	1.05**
		[1.00]	[0.72]	[0.00]	[0.00]	[0.00]	[0.51]	[0.39]	[0.03]
NS1-X	1.05	0.98	1.01**	1.04	1.08	1.10	1.04**	1.05	1.06**
		[0.00]	[0.86]	[0.00]	[0.00]	[0.00]	[0.34]	[0.58]	[0.03]
ATSM-X	1.13	0.85**	1.13*	1.18	1.29	1.42	1.04**	0.99**	1.06**
		[1.00]	[0.87]	[0.00]	[0.00]	[0.00]	[0.19]	[0.06]	[0.00]
Panel C: Forecast combinations									
FC-EW	1.02	0.91	0.98	1.08	1.08	1.09	1.02	1.02	1.03
FC-MSPE	1.02	0.89	0.98	1.07	1.06	1.08	1.02	1.02	1.03
FC-EW-X	1.01	0.85	0.97	1.03	1.06	1.09	1.00	1.01	1.01
FC-MSPE-X	1.00	0.85	0.96	1.01	1.04	1.07	1.00	1.01	1.01
FC-EW-ALL	1.00	0.86	0.94	1.02	1.05	1.08	1.01	1.01	1.02
FC-MSPE-ALL	1.00	0.85	0.94	1.01	1.04	1.07	1.01	1.01	1.02
FC-MCS-EW	0.99	0.82	0.94	0.97	0.98	0.99	1.01	1.01	1.01
FC-MCS-MSPE	0.99	0.82	0.94	0.97	0.98	0.99	1.01	1.01	1.01

Notes: The table reports the [Trace] Root Mean Squared Prediction Error ([T]RMPSE) for *individual* yield models, without and with macro factors in Panels A and B, respectively. Panel C shows results for different forecast *combination* methods. All results are for a 1-month forecast horizon for the out-of-sample period 1994:1 - 2003:12 ($R = 120$ forecasts). The first line in the table reports the value of [T]RMSPE (expressed in basis points) for the Random Walk model (RW), while all other lines reports statistics relative to the RW. Numbers smaller than one (shown in **bold**) indicate that models outperform the random walk, whereas numbers larger than one indicate underperformance. Two stars indicate that a model belongs to the model set $\widehat{\mathcal{M}}_{0.25}^*$ whereas models with one star belong to $\widehat{\mathcal{M}}_{0.10}^*$. The following model abbreviations are used in the table: RW stands for the Random Walk, (V)AR for the first-order (Vector) Autoregressive Model, NS2-(V)AR for the two-step Nelson-Siegel model with a (V)AR specification for the factors, NS1 for the one-step Nelson-Siegel model, ATSM for the affine model. The affix "X" indicates that macro factors have been incorporated in a model as additional explanatory variables. FC-EW and FC-MSPE stand for forecast combinations based on equal weights and MSPE-based weights, respectively, and FC-MCS for forecasting combinations using the pre-filtered model set $\widehat{\mathcal{M}}_{0.25}^*$. For the forecast combinations "X" indicates that forecasts are combined only from models with macro factors whereas "-ALL" indicates that forecasts from all models, both macro as well as yield-only, are combined. No affix in the first two rows of Panel C means that yields-only models are combined. The numbers between parentheses in Panels A and B include the fraction of times a model is included in $\widehat{\mathcal{M}}_{0.25}^*$ for the expanding forecast sample 1994:1 - 2003:12 in the FC-MCS-EW and FC-MCS-MSPE schemes. The $\widehat{\mathcal{M}}_{0.25}^*$ for these forecast combination schemes are determined using an expanding window, with the initial window being 1989:1 - 1993:12.

Table 4: [T]RMSPE 1994:1 - 2003:12, 3-month forecast horizon

Models	[T]RMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	195.81	53.61 [0.00]	48.24** [0.56]	50.71* [0.05]	55.36* [0.14]	59.86* [0.70]	57.25** [1.00]	53.47** [1.00]	49.72** [1.00]
Panel A: Models without macro factors									
AR	1.05	1.11 [0.00]	1.10* [0.00]	1.09* [0.00]	1.08* [0.06]	1.04* [0.61]	1.02** [0.88]	1.03** [0.86]	1.03** [0.98]
VAR	1.10	0.90** [0.41]	1.08* [0.00]	1.21 [0.00]	1.20 [0.00]	1.16 [0.02]	1.09** [0.10]	1.08* [0.44]	1.13** [0.80]
NS2-AR	1.13	1.02* [0.00]	1.16* [0.00]	1.24* [0.00]	1.26 [0.00]	1.23 [0.00]	1.13** [0.03]	1.07** [0.34]	1.06** [0.75]
NS2-VAR	1.05	0.94* [0.11]	0.99** [0.37]	1.08* [0.00]	1.11 [0.00]	1.11* [0.00]	1.06** [0.01]	1.03** [0.64]	1.05** [0.81]
NS1	1.06	1.09 [0.00]	1.09 [0.00]	1.11* [0.00]	1.10* [0.00]	1.10* [0.20]	1.06** [0.56]	1.02** [0.81]	1.03** [0.96]
ATSM	1.06	0.85** [0.68]	0.96** [0.42]	1.11* [0.00]	1.18 [0.00]	1.14 [0.00]	1.02** [0.84]	1.07** [0.00]	1.06** [0.91]
Panel B: Models with macro factors									
AR-X	0.98	0.98 [0.69]	0.95** [1.00]	0.96** [1.00]	0.98** [1.00]	0.98** [1.00]	0.99** [1.00]	0.99** [1.00]	0.99** [1.00]
VAR-X	0.99	0.87** [0.77]	0.98** [0.62]	1.00* [0.00]	1.00* [0.14]	1.03* [0.70]	0.99** [1.00]	0.99** [1.00]	1.00** [1.00]
NS2-AR-X	1.13	1.03 [0.00]	1.24 [0.00]	1.27 [0.00]	1.28 [0.00]	1.20 [0.00]	1.08** [0.57]	1.07** [0.55]	1.04** [0.80]
NS2-VAR-X	1.07	0.85** [0.37]	1.04* [0.00]	1.13* [0.00]	1.19 [0.00]	1.16 [0.00]	1.05** [0.42]	1.05** [0.76]	1.03** [0.92]
NS1-X	1.03	0.84** [0.56]	0.96** [0.62]	1.04* [0.00]	1.10 [0.00]	1.10 [0.02]	1.04** [0.71]	1.03** [0.85]	1.03** [1.00]
ATSM-X	1.04	0.80** [1.00]	0.94** [0.74]	1.04* [0.00]	1.14 [0.00]	1.20 [0.04]	1.03** [0.73]	1.00** [0.86]	1.01** [0.86]
Panel C: Forecast combinations									
FC-EW	1.04	0.94	1.02	1.09	1.10	1.09	1.03	1.02	1.03
FC-MSPE	1.04	0.94	1.02	1.09	1.10	1.09	1.04	1.03	1.04
FC-EW-X	1.00	0.87	0.96	1.01	1.05	1.06	1.00	1.00	0.99
FC-MSPE-X	1.00	0.87	0.95	1.01	1.04	1.05	1.01	1.00	1.00
FC-EW-ALL	1.00	0.86	0.93	1.01	1.05	1.06	1.01	1.00	1.00
FC-MSPE-ALL	1.00	0.86	0.94	1.01	1.05	1.06	1.01	1.01	1.01
FC-MCS-EW	1.00	0.83	0.96	0.97	1.00	1.02	1.01	1.02	1.02
FC-MCS-MSPE	1.00	0.83	0.96	0.97	1.00	1.01	1.01	1.02	1.02

Notes: The table reports forecast results for a 3-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 3 for further details.

Table 5: [T]RMSPE 1994:1 - 2003:12, 6-month forecast horizon

Models	[T]RMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	300.94	83.60*	82.31**	85.20**	89.24**	92.74**	86.36**	79.23**	72.50**
		[0.00]	[0.34]	[0.24]	[0.04]	[0.90]	[0.97]	[0.98]	[1.00]
Panel A: Models without macro factors									
AR	1.07	1.15	1.12**	1.10**	1.10**	1.06**	1.03**	1.04**	1.04**
		[0.00]	[0.00]	[0.09]	[0.02]	[0.67]	[0.92]	[0.86]	[0.90]
VAR	1.20	1.11*	1.22*	1.31*	1.31*	1.24*	1.14**	1.15**	1.21**
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.41]	[0.49]	[0.73]
NS2-AR	1.12	1.05**	1.12**	1.18**	1.22*	1.20*	1.11**	1.06**	1.06**
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.40]	[0.47]	[0.76]
NS2-VAR	1.05	1.02*	1.03**	1.09**	1.11*	1.10**	1.04**	1.02**	1.06**
		[0.00]	[0.12]	[0.09]	[0.00]	[0.04]	[0.55]	[0.93]	[0.90]
NS1	1.06	1.16	1.12**	1.13**	1.11*	1.08**	1.02**	1.00**	1.03**
		[0.00]	[0.00]	[0.00]	[0.00]	[0.15]	[0.70]	[0.85]	[0.80]
ATSM	1.06	0.95**	1.02**	1.12**	1.17*	1.12*	1.01**	1.07**	1.07**
		[0.47]	[0.21]	[0.01]	[0.00]	[0.01]	[0.96]	[0.39]	[0.97]
Panel B: Models with macro factors									
AR-X	1.00	0.97**	0.96**	0.97**	0.99**	1.00**	1.01**	1.00**	1.01**
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
VAR-X	0.98	0.93**	0.98**	1.00**	1.01**	1.00**	0.97**	0.97**	0.99**
		[0.90]	[0.51]	[0.22]	[0.45]	[0.90]	[0.97]	[0.98]	[1.00]
NS2-AR-X	1.13	1.10*	1.22*	1.24*	1.26*	1.18*	1.06**	1.05**	1.04**
		[0.00]	[0.00]	[0.00]	[0.00]	[0.07]	[0.70]	[0.71]	[0.84]
NS2-VAR-X	1.07	0.94**	1.07**	1.14**	1.19*	1.15*	1.04**	1.03**	1.03**
		[0.49]	[0.23]	[0.24]	[0.00]	[0.07]	[0.84]	[0.95]	[0.96]
NS1-X	1.02	0.87**	0.96**	1.03**	1.09*	1.08**	1.01**	1.00**	1.01**
		[0.88]	[0.50]	[0.21]	[0.00]	[0.19]	[0.86]	[0.90]	[0.93]
ATSM-X	1.02	0.84**	0.95**	1.04**	1.11*	1.12**	0.99**	0.98**	1.01**
		[1.00]	[0.61]	[0.24]	[0.00]	[0.32]	[0.97]	[0.85]	[0.88]
Panel C: Forecast combinations									
FC-EW	1.05	1.02	1.05	1.10	1.11	1.08	1.02	1.02	1.04
FC-MSPE	1.05	1.03	1.07	1.10	1.11	1.08	1.02	1.01	1.03
FC-EW-X	0.99	0.90	0.96	1.01	1.04	1.04	0.99	0.98	0.99
FC-MSPE-X	0.97	0.90	0.96	0.99	1.01	1.01	0.96	0.95	0.95
FC-EW-ALL	0.99	0.90	0.95	1.00	1.04	1.03	0.98	0.98	0.99
FC-MSPE-ALL	0.98	0.91	0.96	1.01	1.04	1.02	0.97	0.96	0.97
FC-MCS-EW	0.98	0.92	0.98	0.99	0.98	1.00	0.99	0.99	0.99
FC-MCS-MSPE	0.98	0.91	0.98	0.99	0.98	1.00	0.99	0.99	0.99

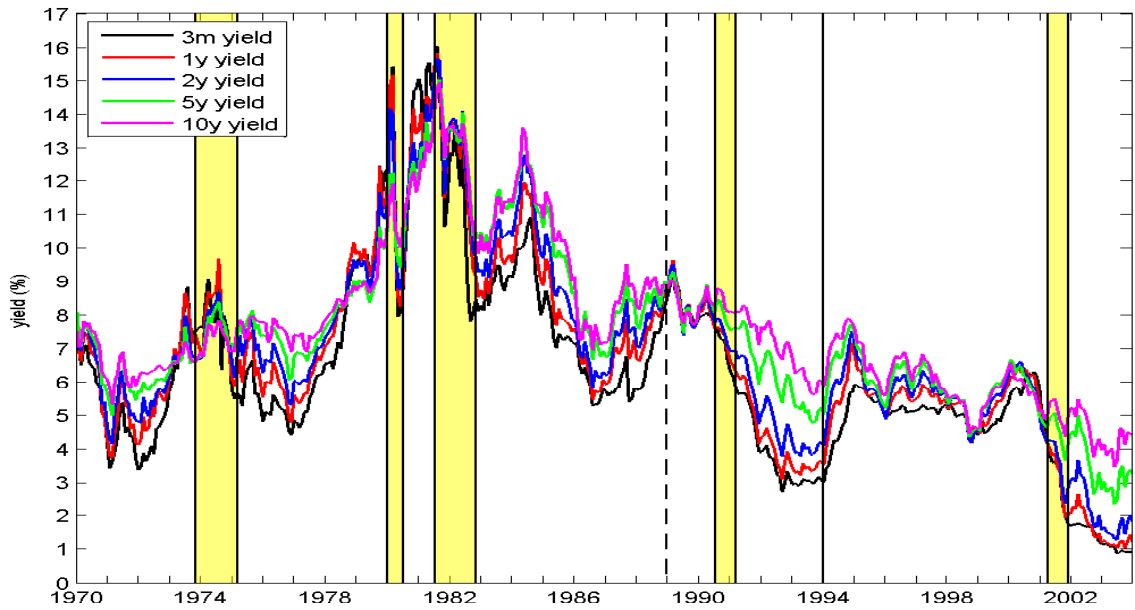
Notes: The table reports forecast results for a 6-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 3 for further details.

Table 6: [T]RMSPE 1994:1 - 2003:12, 12-month forecast horizon

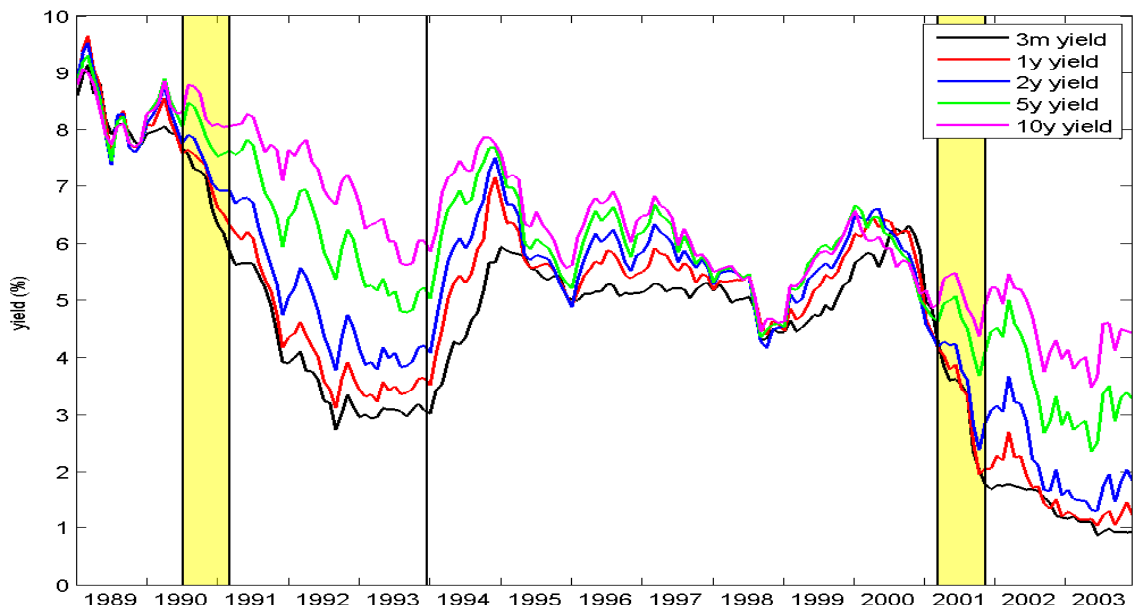
Models	[T]RMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	452.51	136.94** [0.12]	140.61** [0.79]	145.03** [0.94]	146.89** [0.69]	141.77** [0.96]	121.21** [0.95]	108.58** [0.95]	98.96** [0.85]
Panel A: Models without macro factors									
AR	1.10	1.15** [0.01]	1.11** [0.48]	1.09** [0.71]	1.10** [0.48]	1.09** [0.85]	1.07** [0.89]	1.09** [0.48]	1.10** [0.46]
VAR	1.43	1.36 [0.00]	1.41* [0.00]	1.44* [0.01]	1.42* [0.02]	1.40* [0.21]	1.41* [0.28]	1.46 [0.23]	1.55 [0.18]
NS2-AR	1.10	1.02** [0.00]	1.04** [0.35]	1.06** [0.45]	1.10** [0.29]	1.14** [0.36]	1.13** [0.57]	1.12** [0.47]	1.13** [0.68]
NS2-VAR	1.08	1.09** [0.03]	1.07** [0.65]	1.08** [0.70]	1.08** [0.59]	1.09** [0.64]	1.07** [0.77]	1.07** [0.93]	1.12** [0.81]
NS1	1.09	1.21* [0.00]	1.15** [0.00]	1.13** [0.15]	1.10** [0.28]	1.09** [0.48]	1.05** [0.77]	1.04** [0.91]	1.08** [0.81]
ATSM	1.10	1.06** [0.15]	1.07** [0.72]	1.11** [0.35]	1.14** [0.22]	1.12** [0.50]	1.04** [0.93]	1.12** [0.63]	1.13** [0.95]
Panel B: Models with macro factors									
AR-X	1.02	0.95** [1.00]	0.95** [1.00]	0.98** [1.00]	1.00** [1.00]	1.03** [1.00]	1.06** [1.00]	1.05** [1.00]	1.06** [0.92]
VAR-X	0.98	0.97** [0.67]	0.97** [0.92]	0.98** [0.95]	0.99** [0.96]	0.99** [0.96]	0.96** [0.94]	0.97** [0.95]	0.99** [0.95]
NS2-AR-X	1.14	1.15** [0.19]	1.19** [0.50]	1.18** [0.55]	1.19** [0.46]	1.16** [0.46]	1.09** [0.86]	1.09** [0.72]	1.08** [0.76]
NS2-VAR-X	1.11	1.07** [0.60]	1.13** [0.80]	1.14** [0.79]	1.17** [0.58]	1.15** [0.63]	1.07** [0.94]	1.06** [0.95]	1.05** [0.85]
NS1-X	1.01	0.91** [0.60]	0.96** [0.80]	1.00** [0.76]	1.05** [0.52]	1.06** [0.73]	1.01** [0.90]	1.00** [0.93]	1.01** [0.83]
ATSM-X	1.02	0.93** [0.82]	0.99** [0.88]	1.04** [0.74]	1.07** [0.53]	1.08** [0.81]	0.99** [0.94]	1.00** [0.94]	1.02** [0.86]
Panel C: Forecast combinations									
FC-EW	1.08	1.08	1.08	1.09	1.10	1.09	1.07	1.08	1.11
FC-MSPE	1.09	1.11	1.10	1.11	1.11	1.10	1.07	1.08	1.10
FC-EW-X	1.00	0.94	0.97	0.99	1.02	1.03	1.00	1.00	1.00
FC-MSPE-X	0.95	0.95	0.97	0.97	0.98	0.97	0.94	0.94	0.92
FC-EW-ALL	0.99	0.93	0.95	0.97	1.00	1.01	0.99	1.00	1.01
FC-MSPE-ALL	0.98	0.96	0.98	0.99	1.00	1.00	0.97	0.97	0.97
FC-MCS-EW	1.00	0.99	1.02	1.03	1.03	1.02	0.97	0.99	0.99
FC-MCS-MSPE	1.00	0.99	1.02	1.03	1.02	1.01	0.97	0.98	0.99

Notes: The table reports forecast results for a 12-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 3 for further details.

Figure 1: U.S. zero-coupon yields



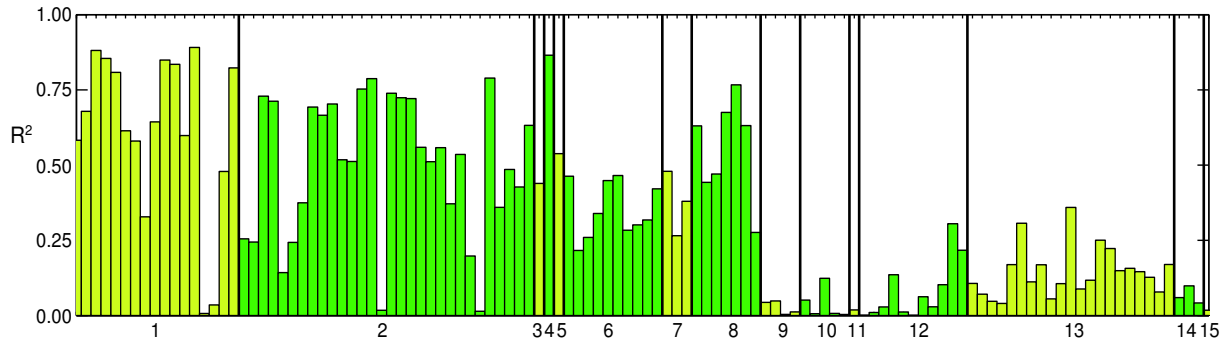
(a) full sample 1970:1 - 2003:12



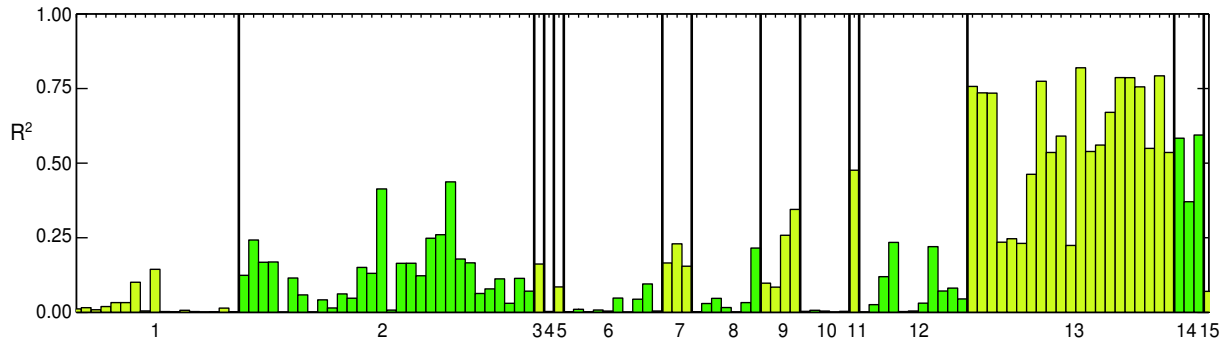
(b) forecast sample 1994:1 - 2003:12

Notes: The figure shows time-series plots of our end-of-month U.S. zero coupon yields (for a selected set of maturities). The yields have been constructed using the Fama and Bliss (1987) bootstrap method. The full sample period is January 1970 - December 2003 (408 observations), and is shown in Panel (a). The solid vertical line shows the beginning of the out-of-sample period January 1994 - December 2003 (120 observations). The start of the initial out-of-sample calibrating period for model weights in the forecast combination scheme is indicated by the dotted line. The calibration and out-of-sample periods are shown separately in Panel (b). Yellow bars highlight NBER recession periods.

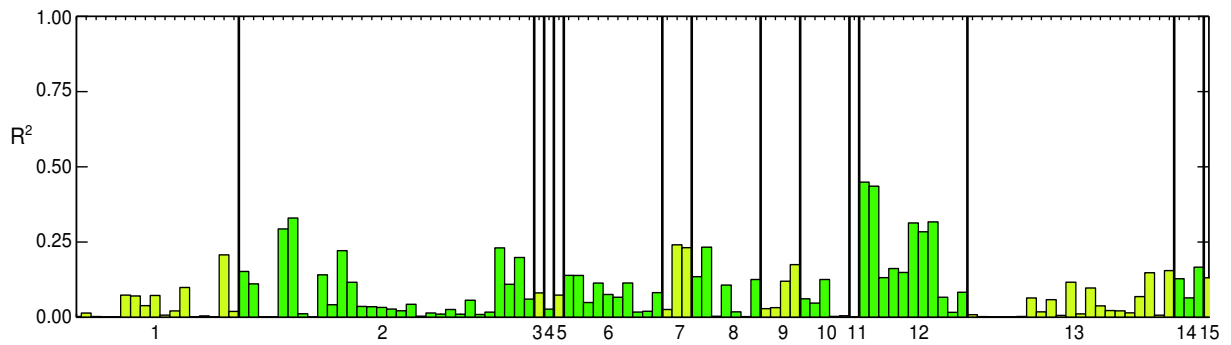
Figure 2: R^2 in regressions of individual macro series on PCA factors



(a) First PCA factor



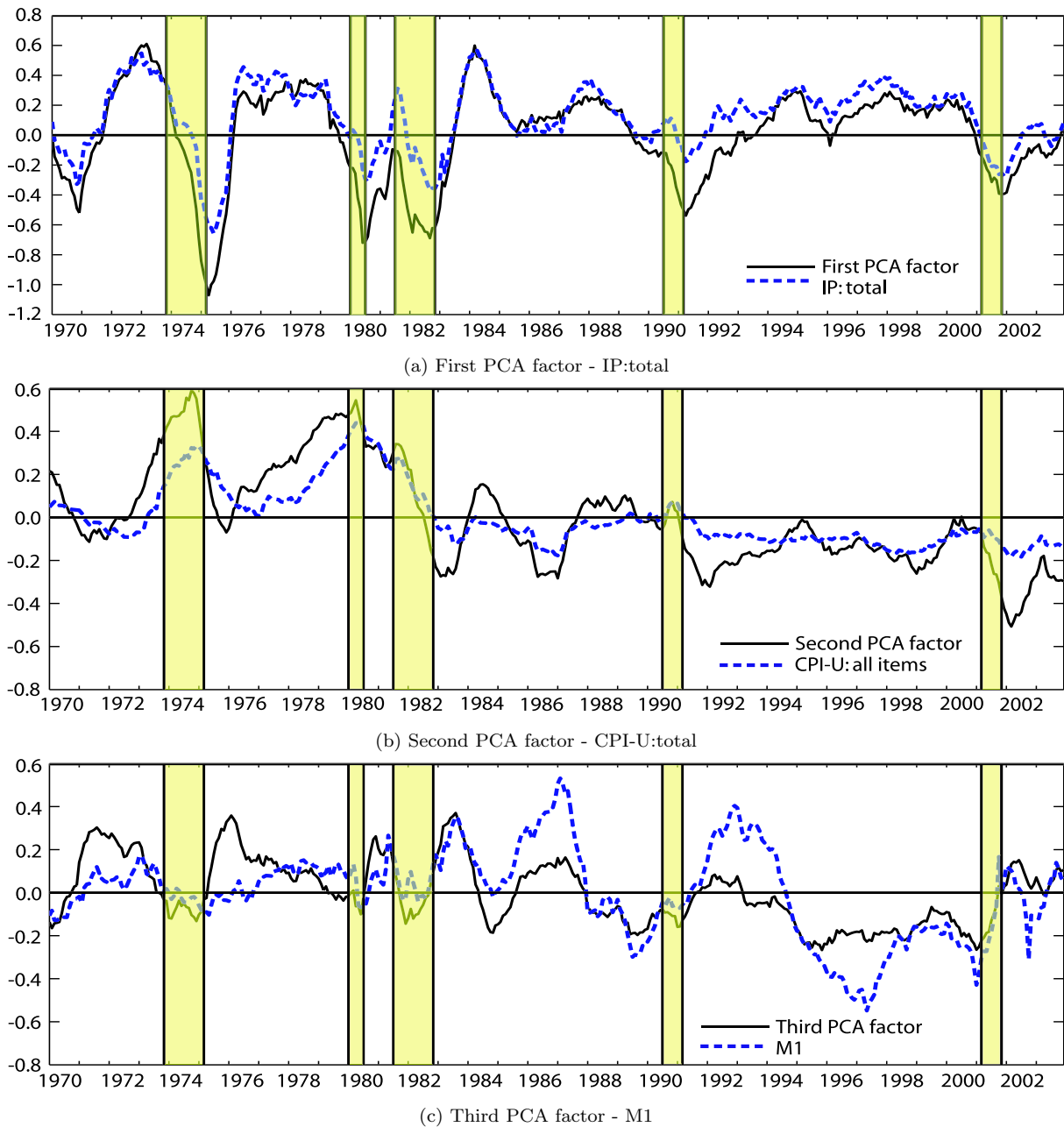
(b) Second PCA factor



(c) Third PCA factor

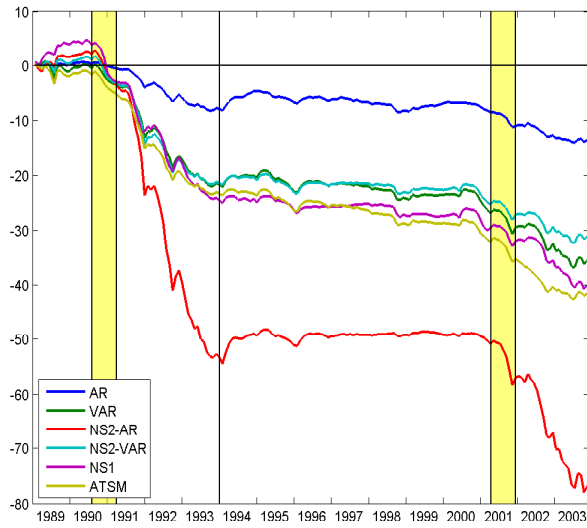
Notes: The figure shows the R^2 when regressing the individual series in the macro panel on each of the first three macro factors. The macro dataset consists of 116 series (transformed to ensure stationarity) and the sample period is January 1970 - December 2003 (408 monthly observations). Panels (a), (b) and (c) show the results for the first, second and third macro factor, respectively. In each panel the macro series are grouped according to the 15 categories as indicated on the horizontal axis. The group categories are (1) real output and income, (2) employment and hours, (3) real retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) prices indices, (14) average hourly earnings and (15) miscellaneous.

Figure 3: Macro factors compared to individual macro series

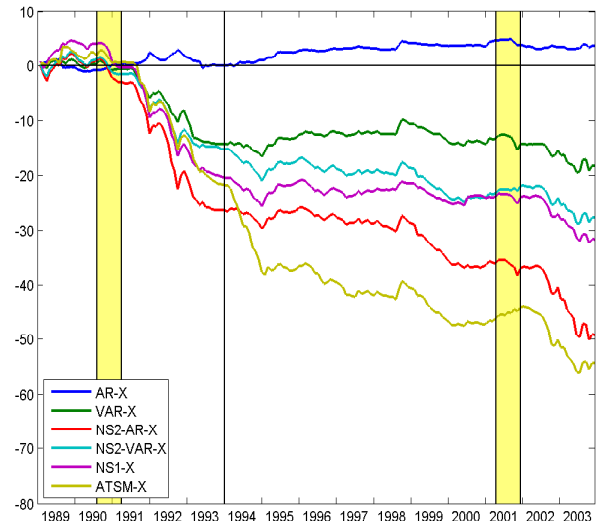


Notes: The figure shows time-series plots of the first three macro factors and the main individual macro series within the category to which the factor is most related. The first factor is plotted together with Industrial Production Index: Total Index (R^2 is 0.88), the second factor is plotted with the Consumer Price Index: All Items (R^2 is 0.77) and the third factor is plotted with Money Stock: M1 (R^2 is 0.44). The macro dataset consists of 116 (transformed to ensure stationarity) series and the sample period used is January 1970 - December 2003 (408 monthly observations). The group categories are (1) real output and income, (2) employment and hours, (3) real retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) prices indices, (14) average hourly earnings and (15) miscellaneous.

Figure 4: Trace Cumulative Squared Prediction Errors, 1-month forecast horizon, 1989 - 2003

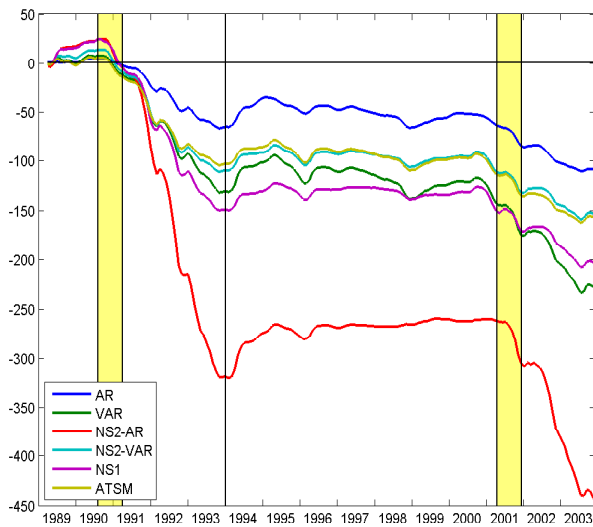


(a) yield-only models

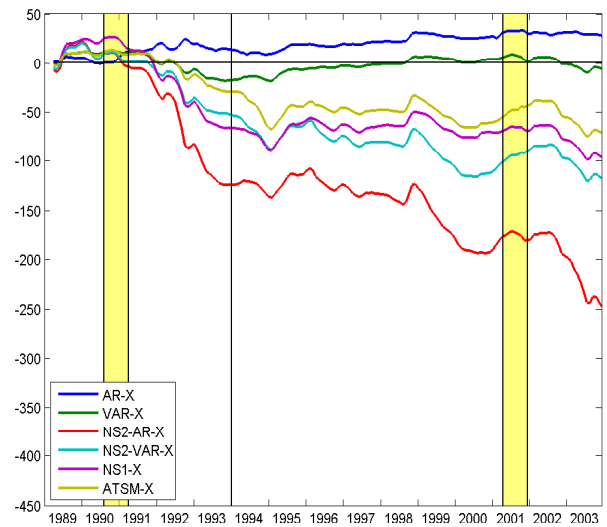


(b) macro models

Figure 5: Trace Cumulative Squared Prediction Errors, 3-month forecast horizon, 1989 - 2003



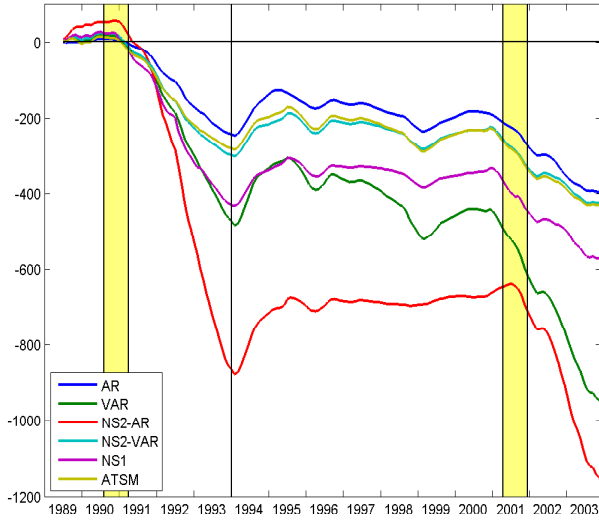
(a) yield-only models



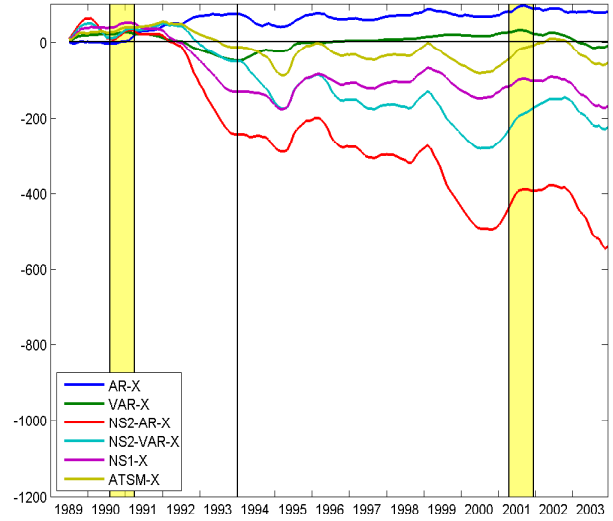
(b) macro models

Notes: Figures 4 and 5 show the Trace Cumulative Squared Prediction Error [TCSPE], relative to the random walk, of individual yield-only models in Panel (a), and individual models with macro factors in Panel (b). Figure 4 shows TCSPEs for a 1-month forecast horizon whereas Figure 5 does so for a 3-month horizon. The forecast sample is 1989:1 - 2003:12.

Figure 6: Trace Cumulative Squared Prediction Errors, 6-month forecast horizon, 1989 - 2003

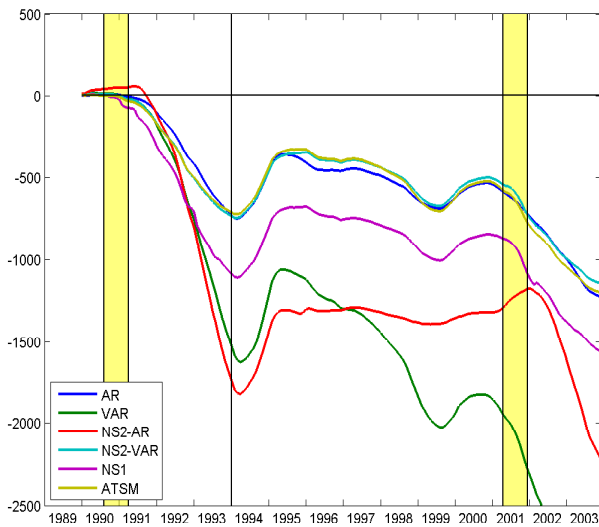


(a) yield-only models

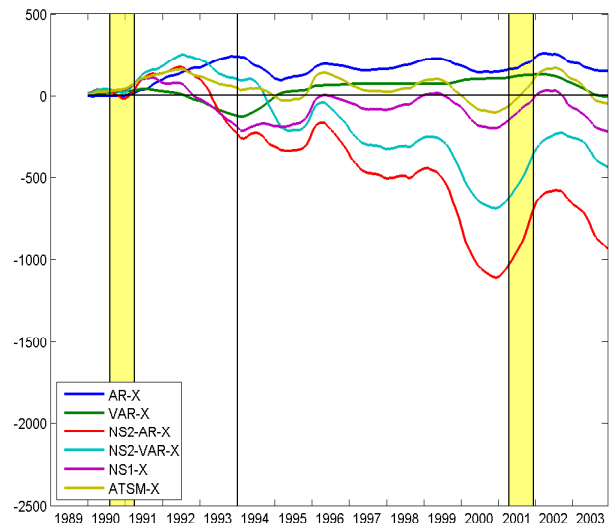


(b) macro models

Figure 7: Trace Cumulative Squared Prediction Errors, 12-month forecast horizon, 1989 - 2003



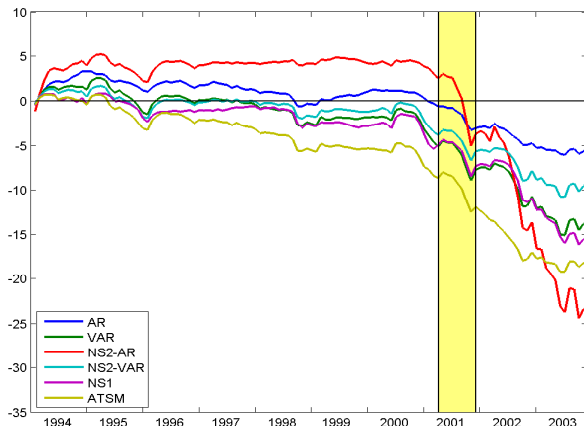
(a) yield-only models



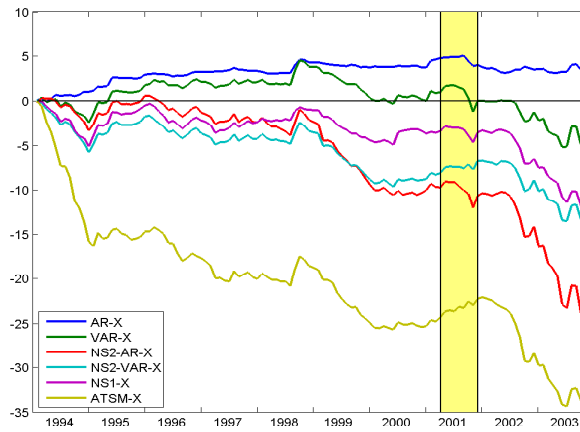
(b) macro models

Notes: Figures 6 and 7 show the Trace Cumulative Squared Prediction Error [TCSPE], relative to the random walk, of individual yield-only models in Panel (a), and individual models with macro factors in Panel (b). Figure 6 shows TCSPEs for a 6-month forecast horizon whereas Figure 7 does so for a 12-month horizon. The forecast sample is 1989:1 - 2003:12.

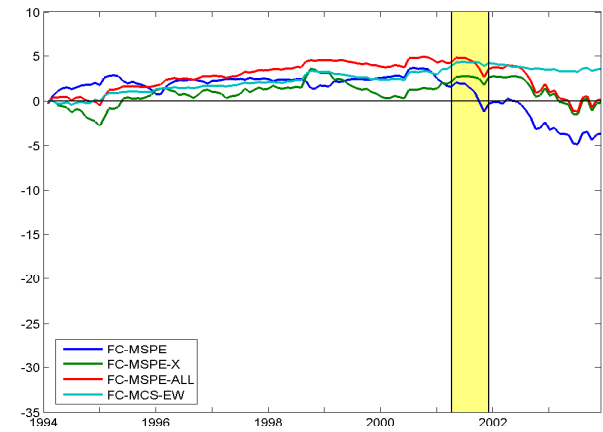
Figure 8: Trace Cumulative Squared Prediction Errors, 1-month forecast horizon, 1994 - 2003



(a) yield-only models

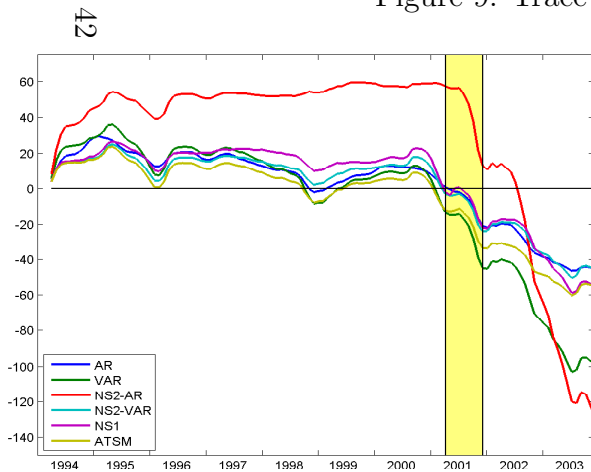


(b) macro models

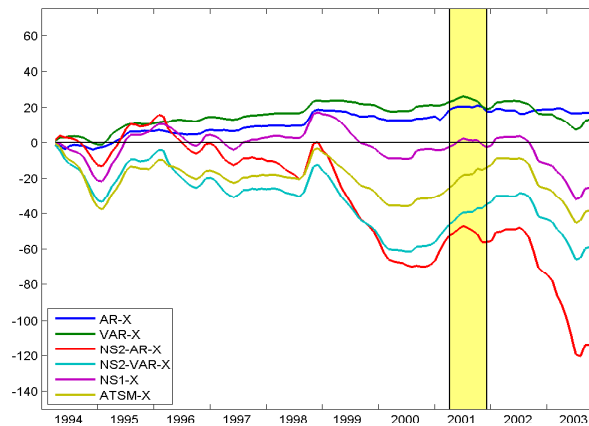


(c) forecast combinations

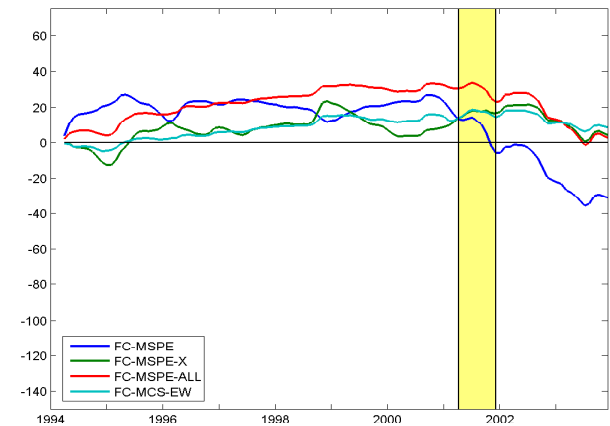
Figure 9: Trace Cumulative Squared Prediction Errors, 3-month forecast horizon, 1994 - 2003



(a) yield-only models



(b) macro models



(c) forecast combinations

Notes: Figures 8 and 9 show the Trace Cumulative Squared Prediction Error [TCSPE], relative to the random walk, of individual yield-only models in Panel (a), individual models with macro factors in Panel (b) and of forecast combinations schemes in Panel (c). Figure 8 shows TCSPEs for a 1-month forecast horizon whereas Figure 9 does so for a 3-month horizon. The forecast sample is 1994:1 - 2003:12.

Figure 10: Trace Cumulative Squared Prediction Errors, 6-month forecast horizon, 1994 - 2003

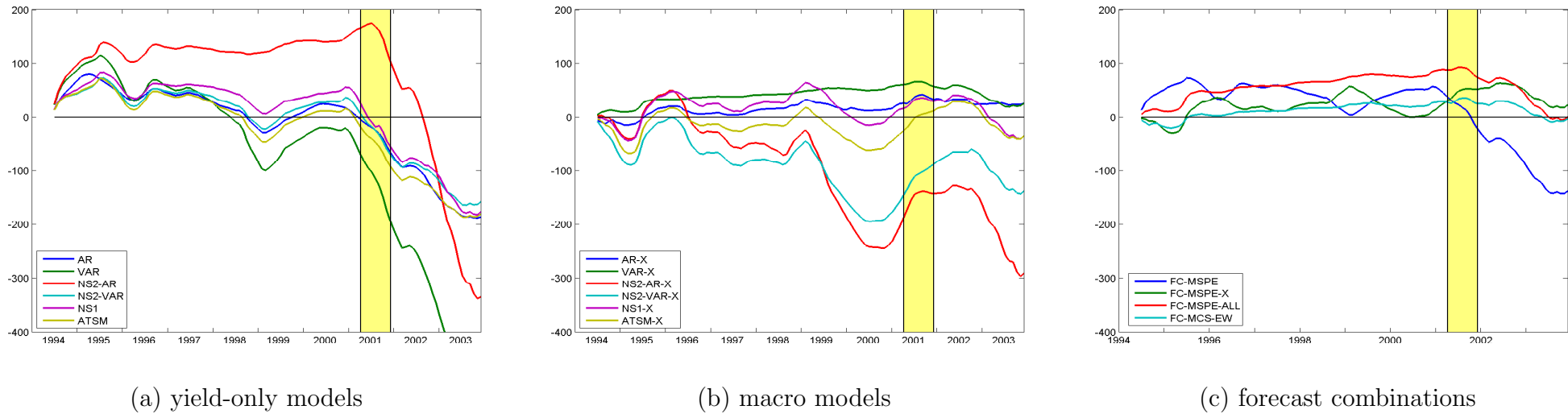
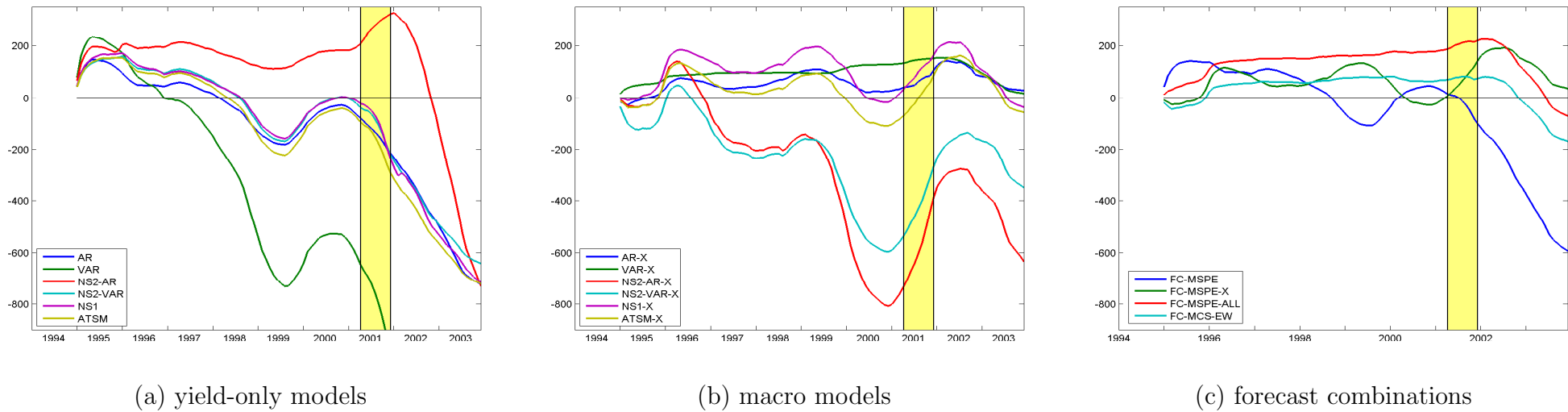
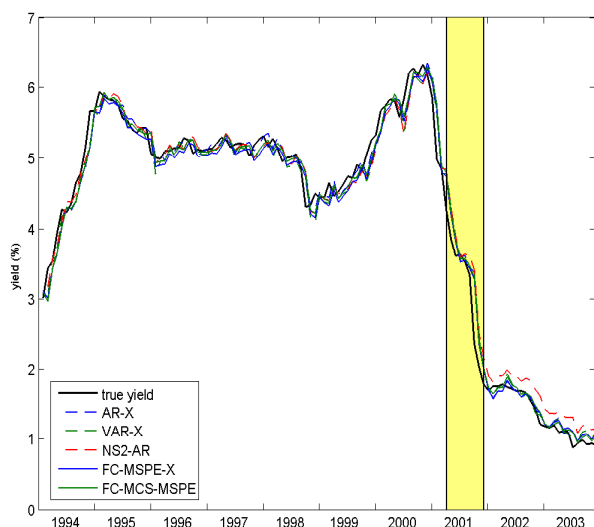


Figure 11: Trace Cumulative Squared Prediction Errors, 12-month forecast horizon, 1994 - 2003

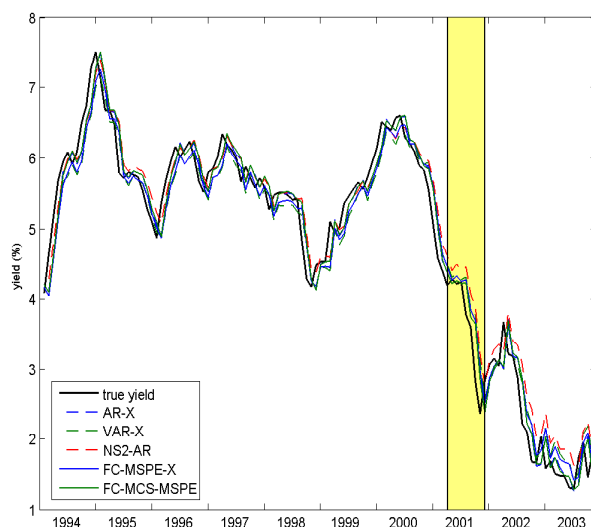


Notes: Figures 10 and 11 show the Trace Cumulative Squared Prediction Error, relative to the random walk, of individual yield-only models in Panel (a), individual models with macro factors in Panel (b) and of forecast combinations schemes in Panel (c). Figure 10 shows TCSPEs for a 6-month forecast horizon whereas Figure 11 does so for a 12-month horizon. The forecast sample is 1994:1 - 2003:12.

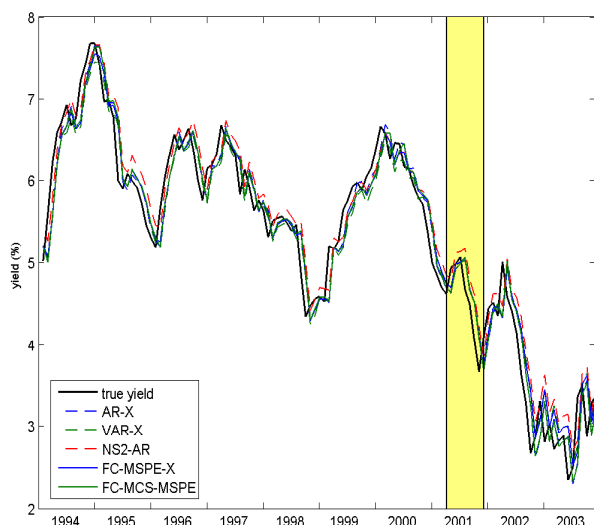
Figure 12: Observed and Predicted Yields, 1-month forecast horizon



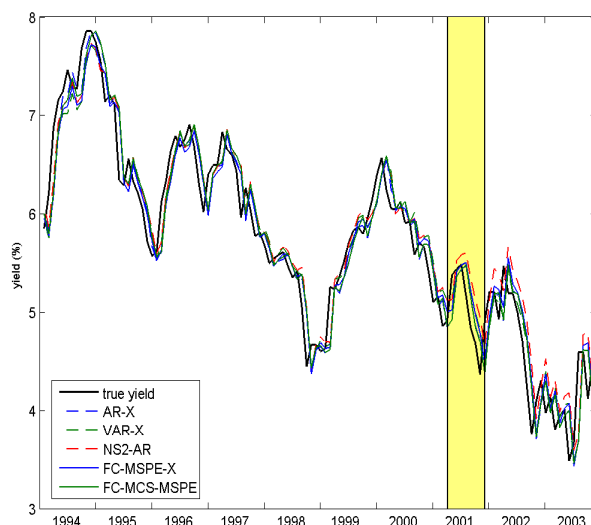
(a) 3-month yield



(b) 2-year yield



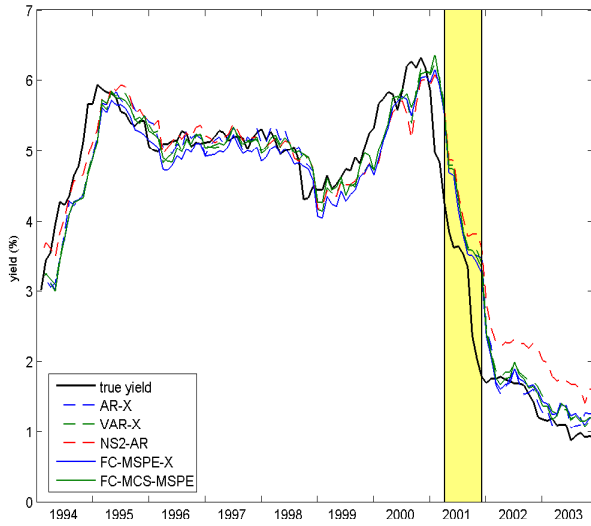
(c) 5-year yield



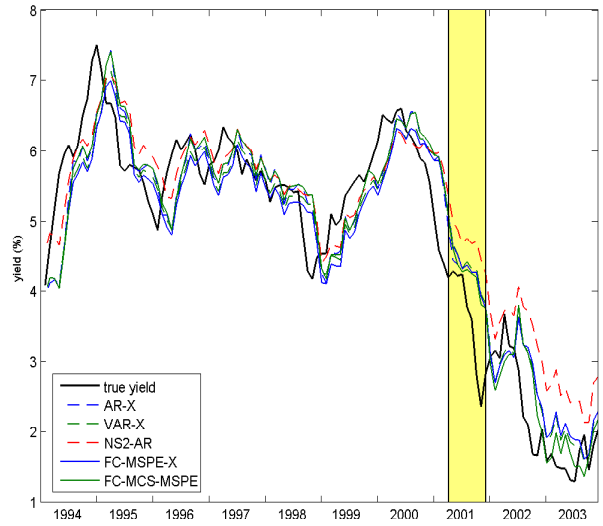
(d) 10-year yield

Notes: The figure shows the observed yields for different maturities (the black solid lines), together with the 1-month forecast from selected models. The dotted lines show forecasts from three individual models: the (Vector) Autoregressive Model with macro factors, and the two-step Nelson-Siegel model (without macro factors). The solid lines are for two forecast combination (FC) schemes: combining models with macro factors using performance based MSPE-weights, and combining model forecasts with MSPE-based weights using only the forecasts from models which are in the Model Confidence Set $\widehat{\mathcal{M}}_{0.25}^*$. Forecasts and observed yields are shown for the out-of-sample period 1994:1 - 2003:12. The forecast are constructed using an expanding estimation window. The out-of-sample period 1989:1 - 1993:12 is used to determine the initial FC weights, and after that an expanding sample is used to compute combination weights.

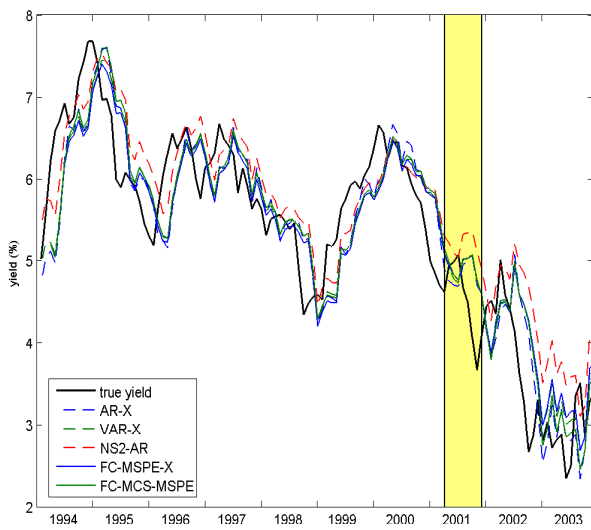
Figure 13: Observed and Predicted Yields, 3-month forecast horizon



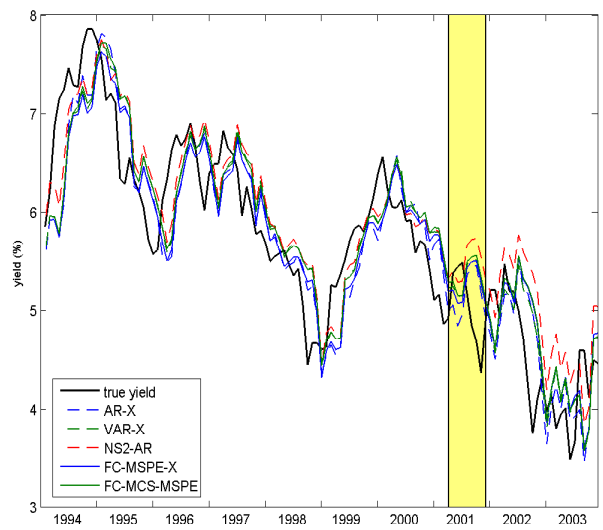
(a) 3-month yield



(b) 2-year yield



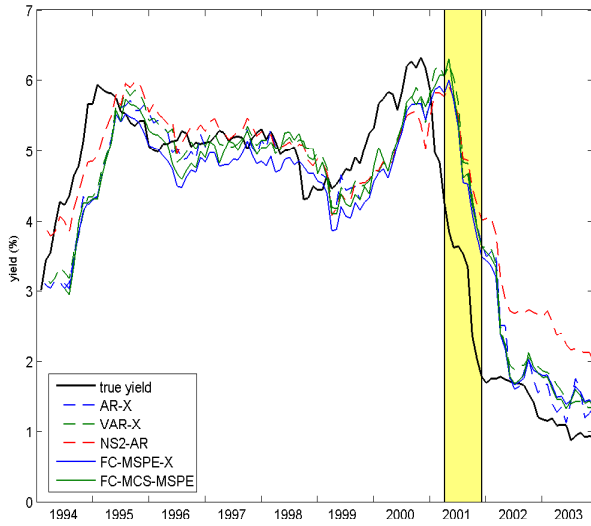
(c) 5-year yield



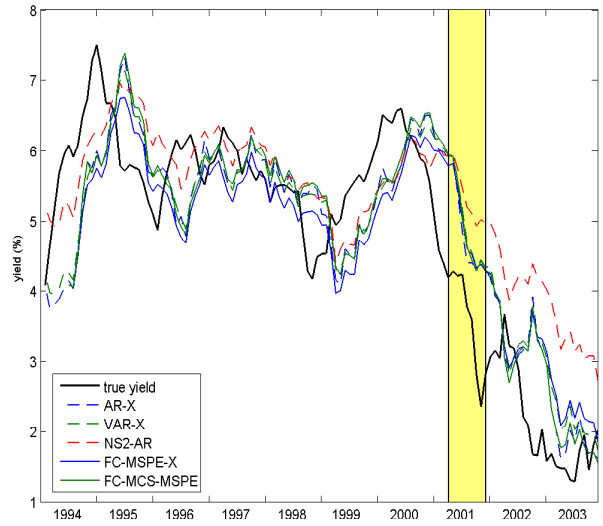
(d) 10-year yield

Notes: The figure shows the observed yields for different maturities, together with the 3-month forecast from selected models. See Figure 12 for further details.

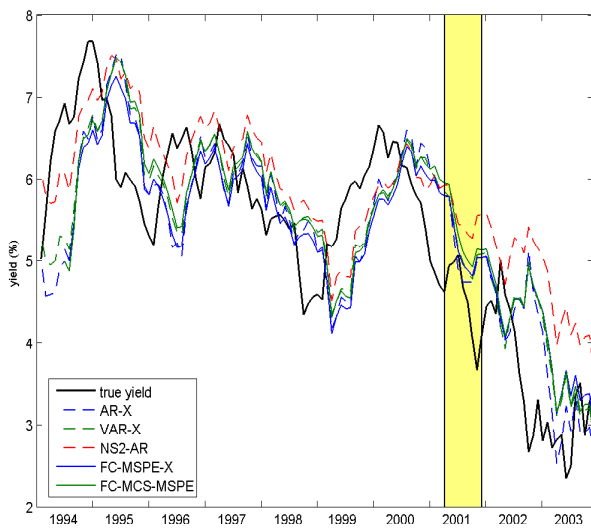
Figure 14: Observed and Predicted Yields, 6-month forecast horizon



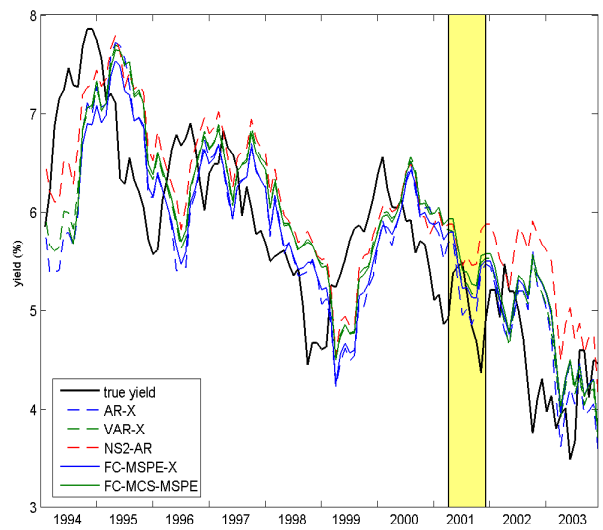
(a) 3-month yield



(b) 2-year yield



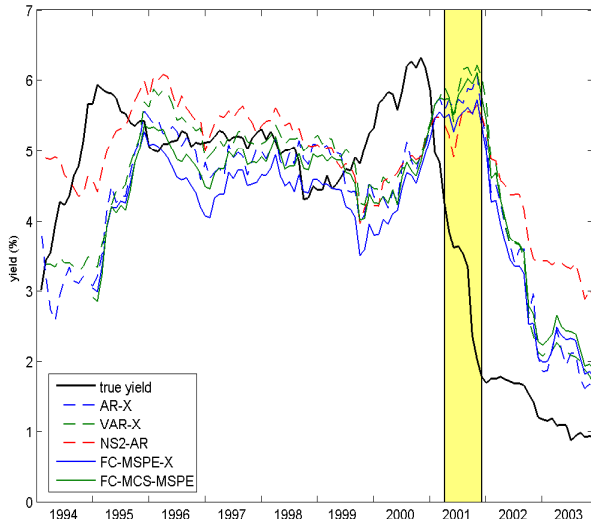
(c) 5-year yield



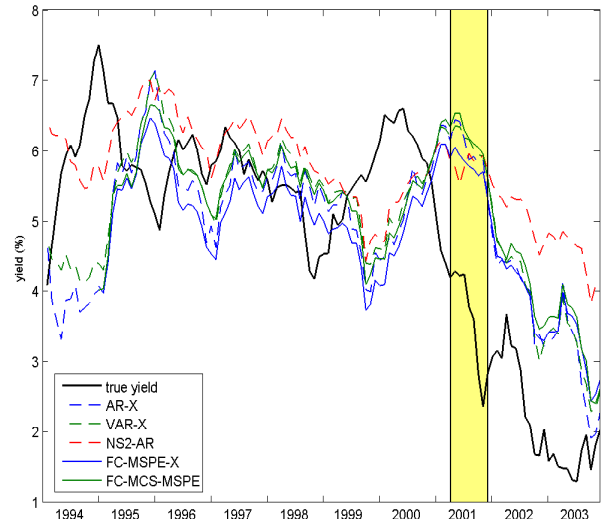
(d) 10-year yield

Notes: The figure shows the observed yields for different maturities, together with the 6-month forecast from selected models. See Figure 12 for further details.

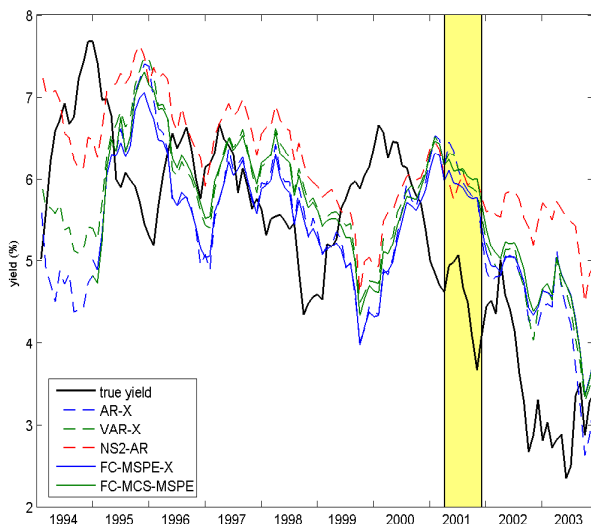
Figure 15: Observed and Predicted Yields, 12-month forecast horizon



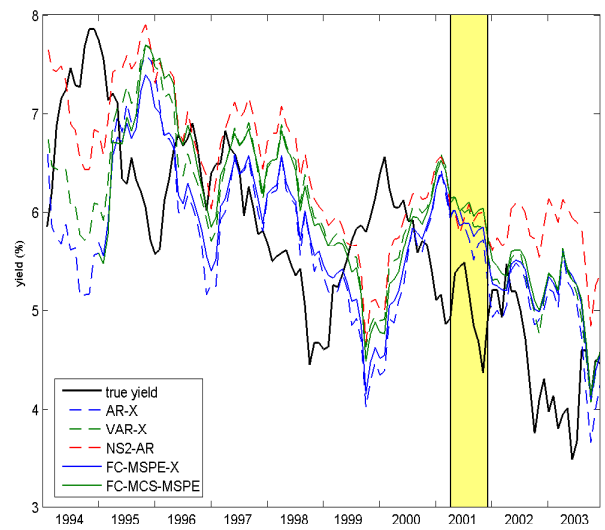
(a) 3-month yield



(b) 2-year yield



(c) 5-year yield



(d) 10-year yield

Notes: The figure shows the observed yields for different maturities, together with the 12-month forecast from selected models. See Figure 12 for further details.