

FORECASTING THE YIELD CURVE WITH DYNAMIC FACTORS

Erhard RESCHENHOFER¹
Thomas STARK²

Abstract

Using two monthly yield datasets over the periods 1970-2000 and 1990-2019, respectively, we re-examine previous findings that yield forecasts based on AR models for the dynamic factors obtained from the Nelson-Siegel curve outperform the random walk forecast and other competitors. Our empirical results do not support these findings. Only the forecasts based on AR models for the differenced yields outperform the random walk forecast. In general, the 1-month-ahead forecasts based on the dynamic factors come out worse than those based on the yields. In the case of 12-months-ahead forecasting, all forecasts perform poorly, particularly those based on AR models fitted to undifferenced time series. Seemingly more positive results obtained in previous studies are explained by a focus on a too short evaluation period.

Keywords: Nelson-Siegel curve, term structure, dynamic factors, out-of-sample forecasting, random walk benchmark, long-term forecasting

JEL Classification: C22, E43

1. Introduction

Identifying and forecasting the factors that drive the term structure of interest rates is of great importance for pricing financial assets, managing portfolios and financial risk, and conducting monetary policy. Numerous models have been proposed for the description of the yield curve. Most of them turned out to be useless for forecasting purposes. However, the class of dynamic factor models seems to constitute an exception. Using a model with three factors, Diebold and Li (2006) obtained forecasts of the yield curve by estimating univariate autoregressive (AR) models for the individual factors as well as vector autoregressive (VAR) models for all three factors. The fact that the AR models fared better than the VAR models (which may depend on the respective forecast period; see Favero *et al.*, 2012) was explained by the fact that the three factors were not highly correlated with each other. Apparently, the forecasts based on the former models were much more accurate

¹ University of Vienna, Department of Statistics and Operations Research, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria. Corresponding author,
E-mail: erhard.reschenhofer@univie.ac.at.

² University of Vienna, Department of Statistics and Operations Research, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria.

at long horizons than various benchmark forecasts based on slope regression, forward rate regression (Fama and Bliss, 1987), forward curve regression (Cochrane and Piazzesi, 2005), principal components, AR models, VAR models, and error correction models (ECMs). While the dynamic factor model used by Diebold and Li (2006) might perform well empirically, it has the theoretical disadvantage that it admits arbitrage possibilities (see Filipović, 1999; for a remedy of this theoretical weakness see Christensen *et al.*, 2011; Niu and Zeng, 2012).

Subsequent papers dealing with out-of-sample forecasting referred to these findings but did not provide any corroborating evidence. Although Matsumura *et al.* (2010) studied a large number of linear term-structure models, none of these models could consistently outperform the random walk model. The inclusion of macroeconomic variables or no-arbitrage restrictions did not improve the out-of-sample fit. Muringi and Kwinjo (2014) compared the 3-factor model only with a 4-factor extension (Svensson, 1994) but did not include standard benchmarks. In Chen and Niu's (2014) study, the factor-forecasting approach (no matter whether AR models or VAR models were used) was inferior to the random walk model. The authors attributed this poor performance to non-stationarities and therefore proposed local AR (LAR) models with time-varying parameters. In addition, they also tried LAR models with exogenous variables (LARX models). Of course, the latter models are not directly comparable to those employed by Diebold and Li (2006).

In order to resolve the discrepancy between the conflicting empirical results, we take a second look at the study of Diebold and Li (2006) and reanalyze their data set. Our focus is on the continuous assessment of the forecasting performance as opposed to the use of a single measure of forecast accuracy such as the mean of the squared forecast errors (MSE) or the mean of the absolute forecast errors (MAE). We also introduce meaningful nontrivial benchmarks. Finally, our findings are checked by extending the study period both into the past and into the future.

The rest of the paper is organized as follows. In the next section, we describe the modeling framework and the data used. We also run some standard analyses which help to understand the nature of the data. In Section 3, we compare the performance of various forecasting methods. Section 4 concludes.

2. Data and Models

Let $P_t(\tau)$ be the price at time t of a bond with par value 1 and residual time to maturity τ (in months). The yield of this bond at time t is defined as the annualized interest rate, *i.e.*

$$y_t(\tau) = \tau^{1/2} \sqrt{\frac{1}{P_t(\tau)}} - 1.$$

For a parsimonious description of the yield curve

$$\tau \rightarrow y_t(\tau), \tau \geq 0,$$

Nelson and Siegel (1987) proposed the model

$$y_t(\tau) = \beta_{1t} + b_{2t} \frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} - \beta_{3t} \exp(-\lambda_t \tau) + u_t(\tau).$$

Since both the estimation and the interpretation of the estimated parameters are compromised by the similarity of the two nontrivial regressors, Diebold and Li (2006) propagated the slightly modified model

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} + \beta_{3t} \left(\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau) \right) + u_t(\tau)$$

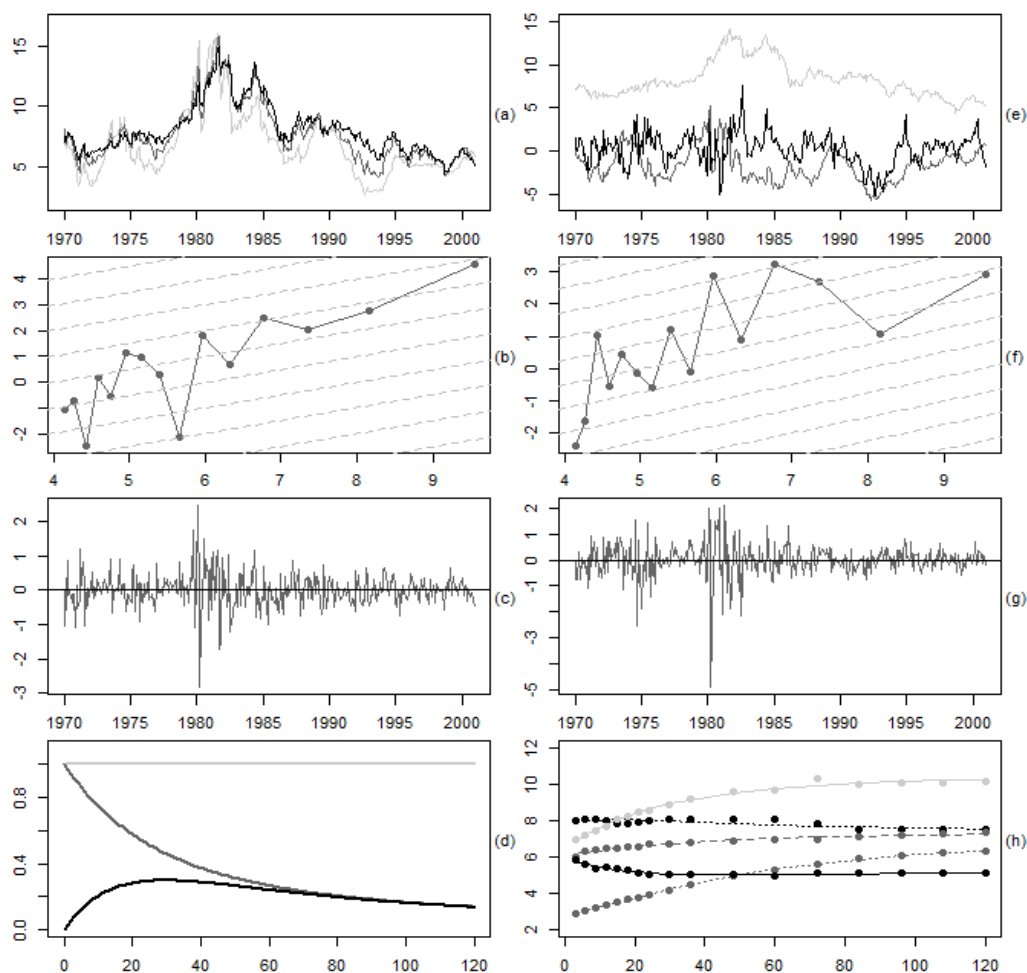
and interpreted the model parameters as dynamic factors representing level, slope and curvature, respectively (for the interpretation of the parameters of similar models fitted to the discount curve rather than to the yield curve see Litzenberger *et al.*, 1995). The loading on the first factor is a constant and therefore does not decrease as τ increases. The loading on the second factor starts at the maximum and then decays quickly towards zero. The loading on the third factor increases from 0 to a maximum and then decays to zero (see Figure 1.d). The factors are therefore also called long-term, short-term, and medium-term, respectively.

As pointed out by Chen and Niu (2014), under a fixed λ_t , any non-stationarity in the yields can be solely attributed to non-stationarities in the factors. Diebold and Li (2006) set $\lambda_t = \lambda_0 = 0.0609$ for all t . This value is commonly thought to ensure that the loading on the medium-term factor achieves its maximum at 30 months (average of two and three-year maturities). Clearly, it does not really matter that the value of λ_t implying a maximum at 30 months is actually given by 0.0598 rather than 0.0609 because the maturity of 30 months was set arbitrarily in the first place. A different choice of the maturity would have a much greater effect. For example, we must set $\lambda_t = 0.074720$ and $\lambda_t = 0.049813$ for a maximum at 24 months and 36 months, respectively. Alternatively, a fixed λ_t can also be chosen by optimizing the fit for the data set used. However, if the first forecast is made in the $m+1^{\text{st}}$ month, only the first m months can be used for the optimization in order to avoid the danger of a data-snooping bias. For the data set and $m=108$ used by Diebold and Li (2006), we obtain 0.06174 and 0.06569 by minimizing the MSE and the MAE, respectively. These values are relatively close to the optimal values 0.06553 and 0.06867 for the whole sample. All computations were carried out with the free statistical software R (R Core Team, 2017).

Diebold and Li (2006) used monthly yield data (obtained from end-of-month price quotes for U.S. Treasuries) from January 1985 to December 2000 ($n=192$) at maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months (for a study on daily data see Sambasivan and Das, 2017; for studies on international data see Kaya, 2013; Shang and Zheng, 2018). In the following, we use a much longer time series ($n=372$) beginning in January 1970 and extending through December 2000 (thankfully made available by Francis X. Diebold on his homepage). Given the $K=17$ time series of yields at different maturities, estimates $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$, $\hat{\beta}_{3t}$ of the time-varying model parameters β_{1t} , β_{2t} , and β_{3t} can, for each t , be obtained by OLS when λ_t is fixed. Observing that the 3-factor model (with $\lambda_t=0.0609$) is capable of replicating a variety of yield curve shapes (see also Figure 1.h),

Figure 1

Analysis of Yields and Factors



Notes:

- (a) Yields at maturities of 3 (gray), 36 (darkgray), and 120 (black) months
- (b) Log periodogram of 36-m yield at first 15 frequencies ω_j plotted against $-2 \log(\sin(\omega_j/2))$
- (c) Changes of 36-m yield
- (d) Factor loadings: long-term (gray), short-term (darkgray), medium-term (black)
- (e) Factors representing level (gray), slope (darkgray) and curvature (black)
- (f) Log periodogram of slope at first 15 Fourier frequencies ω_j plotted against $-2 \log(\sin(\omega_j/2))$
- (g) Changes of slope
- (h) Yield curves (lines) fitted to yields (points) at different times: month=1 (black, dotted), n/4 (darkgray, dashed), n/2 (gray, solid), 3n/4 (darkgray, dotted), n (black, solid)

Diebold and Li (2006) claimed that forecasting the yield curve is equivalent to forecasting the three factors. Using augmented Dickey-Fuller tests, they found indications of unit roots in the factors. This finding can be confirmed by a more robust analysis in the frequency

domain. A steep increase of the periodogram in the neighborhood of frequency zero is a possible indication of a pole at frequency zero. The integrability of the spectral density, which is required for stationarity, depends on the steepness of the increase. Assuming that the spectral density in the neighborhood of frequency 0 can be approximated by

$$f(\omega) \sim C |1 - e^{-i\omega}|^{-2d} = 4^{-d} C \sin^{-2d}\left(\frac{\omega}{2}\right)$$

and

$$\log f(\omega) \sim \log 4^{-d} C + d(-2 \log(\sin(\frac{\omega}{2}))),$$

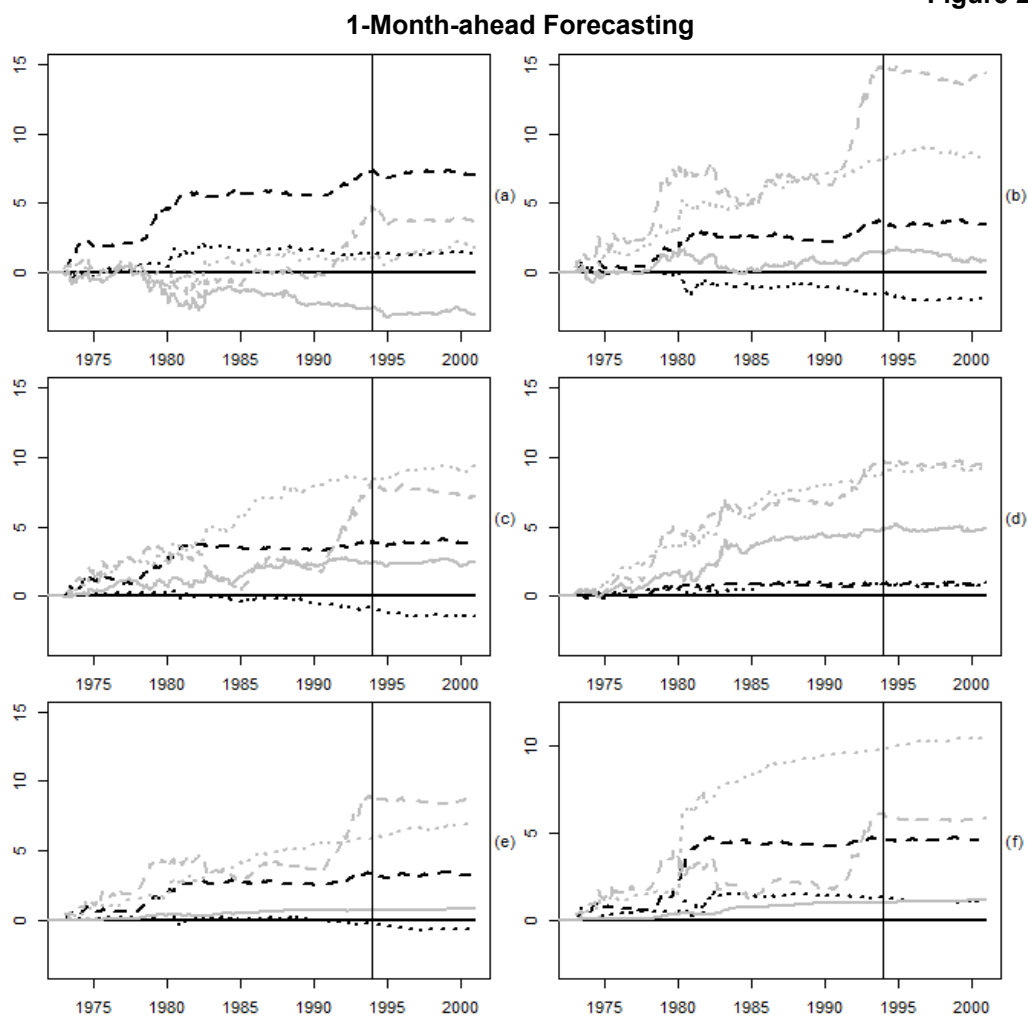
respectively, and plotting the first log periodogram ordinates $\log(f(\omega))$ against $-2 \cdot \log(\sin(\omega/2))$, we can check whether the slope d is greater than 0.5, which is inconsistent with stationarity. Figures 1.b (obtained from the 36 months yields) and 1.f (obtained from the second factor) suggest that the factors and the yields are indeed non-stationary.

Many economic and financial time series are non-stationary before differencing and do not have large autocorrelations after differencing. Usually, good forecasts are obtained by fitting low-order AR models to the differences. The order 1 is often the best choice. However, Diebold and Li (2006) use only the random walk model, which corresponds to an AR model of order 0 for the differences. They consider nontrivial AR models for the yield and factor levels, VAR models for the yield and factor levels as well as for the yield changes but take a pass on the most promising models, namely AR models for the yield and factor changes (depicted in Figures 1.c and 1.g). Based on ample empirical evidence from the macroeconomics literature, one might expect that these omitted models have, a priori, the best chance to actually outperform the random walk model. They are therefore included in our study on yield-curve forecasting, the results of which will be presented in the next section.

3. Out-of-Sample Forecasting

In this section, we compare the performance of various term-structure forecasts for the forecast horizons of $h=1$ and $h=12$ months. We forecast recursively with expanding window length $m \geq 36$, hence the data from January 1970 to December 1972 are used to forecast the yields in January 1988 and the data from January 1970 to November 2000 are used to forecast the yields in December 2000. Figure 2 shows the cumulative absolute 1-step-ahead forecast errors for the maturities of 3 (2.a), 12 (2.b), 36 (2.c), and 120 (2.d) months relative to the random-walk benchmark. Averages over all 17 maturities are shown in 2.e. Since the cumulative squared 1-step-ahead forecast errors are essentially comparable (but are clearly more erratic due to their sensitivity to outliers), only the averages over all 17 maturities are shown in 2.f. Overall, only the forecasts based on the differenced yields are slightly better than the benchmark. Apparently, the factors are not helping at all. The forecast based on the last yield performs better than the forecast based on the most recently fitted yield curve. Similarly, the forecasts based on AR(1) models for the yields perform better than those based on AR(1) models for the factors. Finally, forecasts based on AR(1) models for the differenced yields outperform forecasts based on AR(1) models for the differenced factors. In view of the non-stationarities in both the yields and the factors, it is not surprising that differencing has, in general, a positive effect.

Figure 2



Notes: Cumulative absolute errors of 1-month-ahead forecasts of monthly yields at different maturities relative to random-walk benchmark (a: 3 months, b: 12, c: 36, d: 120, e: average over all 17 maturities; f: average over all 17 maturities for cumulative squared errors; black vertical line: start of forecasting period of Diebold and Li, 2006).

- Black, solid: Random-walk forecasts based on yields
- Black, dashed: Forecasts based on AR(1) models for the yields
- Black, dotted: Forecasts based on AR(1) models for the differenced yields
- Gray, solid: Random-walk forecasts based on fitted 3-factor model
- Gray, dashed: Forecasts based on AR(1) models for the factors
- Gray, dotted: Forecasts based on AR(1) models for the differenced factors

The Diebold-Mariano (1995) test can be used to test the null hypothesis that two forecasts have the same accuracy. In our case, the results of pairwise testing confirmed that only the

forecasts based on AR(1) models for the differenced yields can keep up with the benchmark. The details are given in Table 1. The results obtained with the modified Diebold-Mariano test (Harvey *et al.*, 1997) are not reported because they are practically identical. The actual testing was carried out with the help of the function DM.test of the R package multDM (in contrast to Table 1, the specifications “more” and “less” in this function refer to the alternative hypothesis rather than to the null hypothesis).

Table 1
Diebold-Mariano test (5%) based on Squared/Absolute Forecasting Errors:
Number of Rejections for 17 Maturities of the Hypothesis that Forecasts are
More/Less Accurate than Benchmark (Random Walk for Yields)

Forecasts	SE “less”	AE “less”	SE “more”	AE “more”
AR(1) model for the yields	0	0	17	14
AR(1) model for the yield changes	0	4	0	0
Random walk for the factors	0	0	3	3
AR(1) model for the factors	0	0	8	14
AR(1) model for the factor changes	0	0	15	15

Given the disappointing performance of the 1-step forecasts, we do not expect that long-term forecasts are of any use. However, encouraged by the findings of Diebold and Li (2006), we still take a closer look at the forecast horizon of $h=12$ months.

The coefficients \hat{c}_{jt} , $\hat{\gamma}_{jt}$ and \hat{d}_{jt} , $\hat{\delta}_{jt}$ in the yield forecasts

$$\begin{aligned} \hat{y}_{t+h}(\tau) &= (\hat{c}_{1t} + \hat{\gamma}_{1t}\hat{\beta}_{1t}) + (\hat{c}_{2t} + \hat{\gamma}_{2t}\hat{\beta}_{2t}) \underbrace{\frac{1 - \exp(-\lambda_0\tau)}{\lambda_0\tau}}_{=L_2(\tau)} \\ &+ (\hat{c}_{3t} + \hat{\gamma}_{3t}\hat{\beta}_{3t}) \underbrace{\left(\frac{1 - \exp(-\lambda_0\tau)}{\lambda_0\tau} - \exp(-\lambda_0\tau) \right)}_{=L_3(\tau)} + u_t(\tau) \end{aligned}$$

and

$$\begin{aligned} \hat{y}_{t+h}^\Delta(\tau) &= (\hat{\beta}_{1t} + \hat{d}_{1t} + \hat{\delta}_{1t}\Delta\hat{\beta}_{1t}) + (\hat{\beta}_{2t} + \hat{d}_{2t} + \hat{\delta}_{2t}\Delta\hat{\beta}_{2t})L_2(\tau) \\ &+ (\hat{\beta}_{3t} + \hat{d}_{3t} + \hat{\delta}_{3t}\Delta\hat{\beta}_{3t})L_3(\tau) + u_t(\tau) \end{aligned}$$

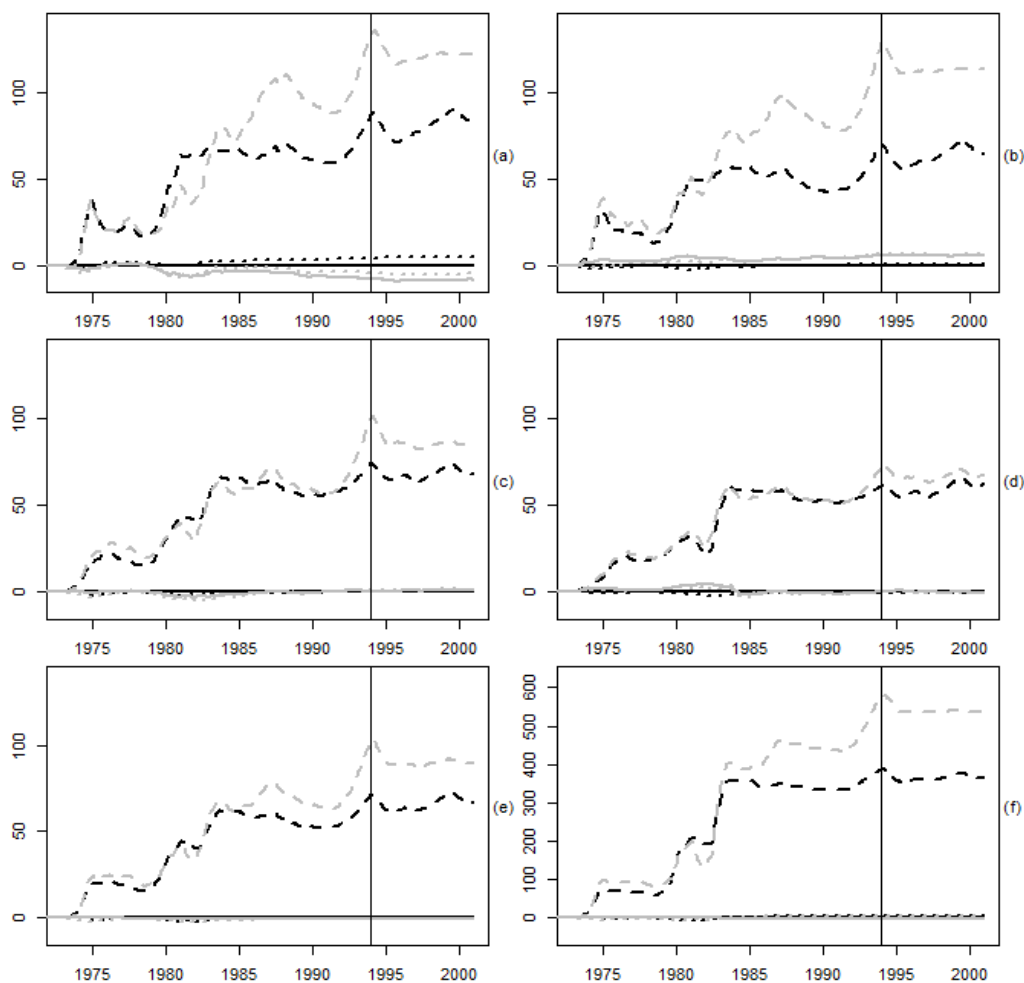
are obtained by regressing $\hat{\beta}_{js}$ on $\hat{\beta}_{j(s-h)}$, $s=h+1, \dots, t$, and $\Delta\hat{\beta}_{js}$ on $\Delta\hat{\beta}_{j(s-h)}$, $s=h+2, \dots, t$, respectively. Analogously, the more straightforward yield forecasts $\check{y}_{t+h}(\tau)$ and $\check{y}_{t+h}^\Delta(\tau)$ based on autoregressive models for the yield levels and the yield changes, respectively, are defined.

Figure 3 shows the cumulative absolute 12-step-ahead forecast errors for different maturities as well as averages over all 17 maturities. Again, the cumulative squared forecast errors are only shown for the averages. There are no forecasts that consistently outperform the random walk benchmark. The forecasts based on autoregressive models for the levels come out worst. However, Figure 3 also shows that it is possible to find subperiods where these forecasts slightly outperform the benchmark. The longest of these subperiods starts in early

1984. Luckily, Diebold and Li (2006) chose just that starting point for their comparison of the different forecasts.

Figure 3

12-Months-ahead Forecasting



Notes: Cumulative absolute errors of 12-months-ahead forecasts of monthly yields at different maturities relative to random-walk benchmark (a: 3 months, b: 12, c: 36, d: 120, e: average over all 17 maturities; f: average over all 17 maturities for cumulative squared errors; black vertical line: start of forecasting period of Diebold and Li, 2006).

Black, solid: Random-walk forecasts based on yields

Black, dashed: Forecasts based on AR(1) models for the yields

Black, dotted: Forecasts based on AR(1) models for the differenced yields

Gray, solid: Random-walk forecasts based on fitted 3-factor model

Gray, dashed: Forecasts based on AR(1) models for the factors

Gray, dotted: Forecasts based on AR(1) models for the differenced factors

4. Conclusions

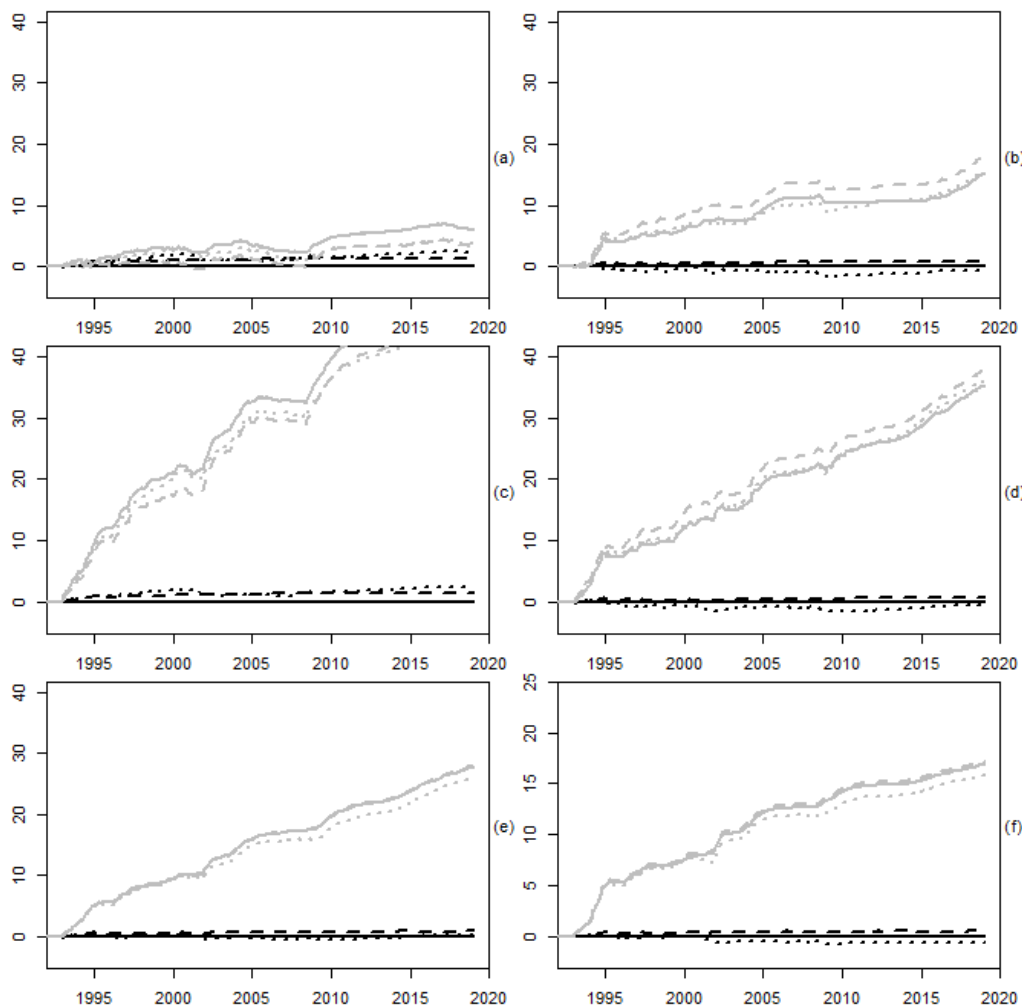
In general, the approximation of individual yields by a smooth parametric function has two effects. A bias is introduced and the variance is reduced. The usefulness of the approximation depends on the sizes of these effects. In the simplest case of random walk forecasting, there are two forecasts that must be compared. The first forecasts a future yield simply by the present yield and the second forecasts a future yield by the approximated present yield. In our empirical study of monthly yields from January 1970 to December 2000, we find that the use of the approximation reduces forecast accuracy indicating that the bias effect outweighs the variance effect. Similarly, when we apply autoregressive models to the levels and the differences of the yields and the factors (the time-changing parameters of the approximation), respectively, we always find that the forecasts based on the yields are better than those based on the factors. Overall, only the forecasts based on autoregressive models for the yield changes slightly outperformed the benchmark (random walk for yields). This outperformance appears to be genuine because it is corroborated by some significant results obtained with the Diebold-Mariano (1995) test.

Of course, there is still room for further development of the factor approach. There are numerous possibilities. But in view of the relatively small sample size, any such attempt would quite rightly raise the suspicion of data snooping. When all the basic variants of factor forecasting appear to be completely useless, we should possibly leave it at that. Moreover, when the quality of short-term forecasts is poor, it does not make sense to bother with long-term forecasting. A comparison of several bad long-term forecasts will most likely result in a random outcome. Just by chance, each forecast will be much better than its competitors in certain subperiods and much worse in others. This is also the case for the forecast propagated by Diebold and Li (2006). It performs very poorly before 1994 and between 1995 and 1999. However, some lucky hits in 1994 and in 2000 suffice to squeeze out a narrow lead over the benchmark in the subperiod from 1994 to 2000, which has been used by Diebold and Li. However, already a slight extension of this subperiod into the past causes a dramatic deterioration in the performance of their forecast.

To find out what happens when we extend the forecasting period into the future, we downloaded a newer dataset (with the help of the R package Quandl) beginning in January 1990 and extending through January 2019. This dataset includes daily yields at maturities of 1, 2, 3, 6, 12, 24, 36, 60, 84, 120, 240, and 480 months. After omitting the two shortest and the two longest maturities (because of too many missing values), we transformed the eight remaining daily yield series to monthly series by retaining only the last quotes of each month. Compared to the dataset studied above, the number of maturities was practically halved. However, the possible negative effects of this reduction on the estimation of the yield curves are dampened by the smoothness of the yield curves and the fact that the range (3-120 months) has remained unchanged. Figure 4, which is analogous to Figure 2, shows first that the forecasts based on AR(1) models for the differenced yields are again the only ones that outperform the benchmark and second that the forecasts based on factors have further fallen behind those based on yields, which is not surprising in view of the fact that the number of maturities has dropped considerably. The implications of this fact are less severe for long-term forecasts, which are very erratic anyway. Indeed, Figure 5, which is analogous to Figure 3, shows that the forecasts based on AR(1) models for the factors still perform similarly to those based on AR(1) models for the yields, which are clearly not affected by the total number of maturities. Apart from short subperiods, both forecasts come off very badly. The financial crisis does not seem to have an important effect on their performance.

Figure 4

1-Month-ahead Forecasting (New Data)



Notes: Cumulative absolute errors of 1-month-ahead forecasts of monthly yields at different maturities relative to random-walk benchmark (a: 3 months, b: 12, c: 36, d: 120, e: average over all 17 maturities; f: average over all 17 maturities for cumulative squared errors; black vertical line: start of forecasting period of Diebold and Li, 2006).

Black, solid: Random-walk forecasts based on yields

Black, dashed: Forecasts based on AR(1) models for the yields

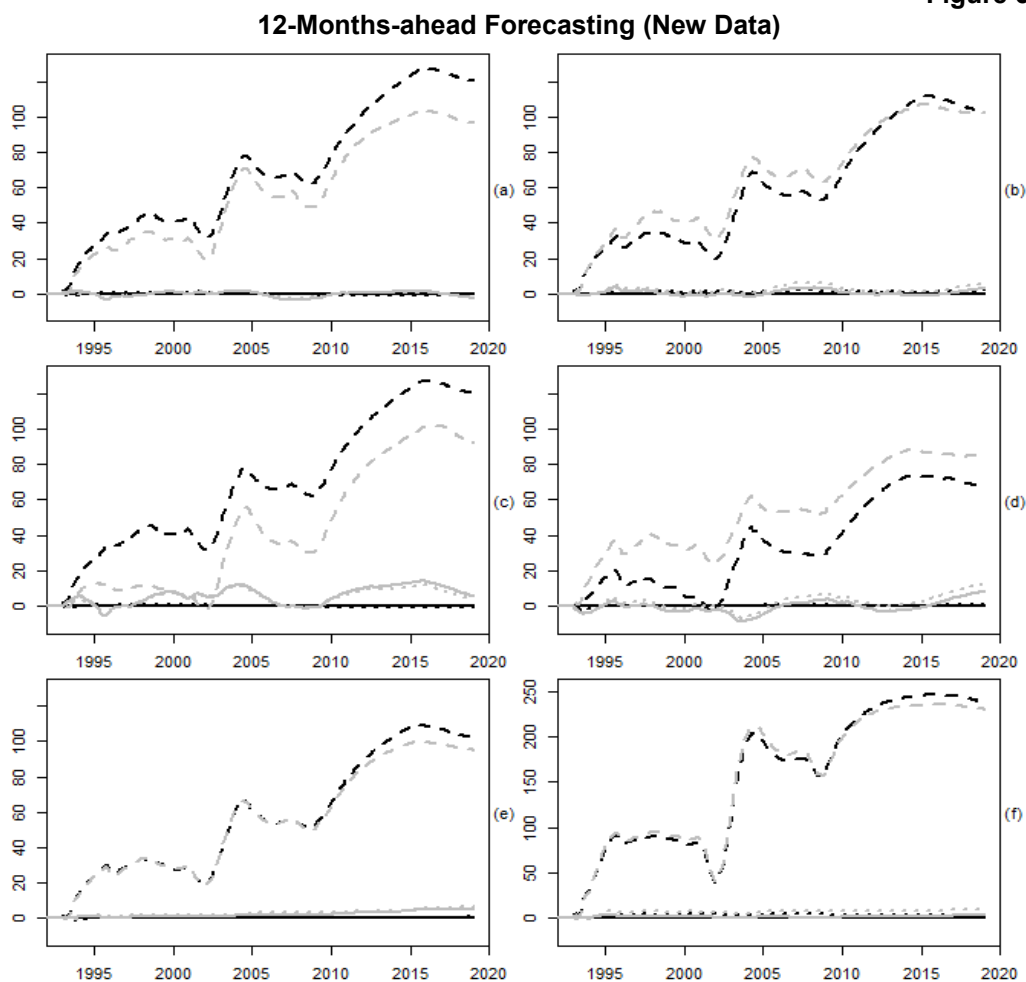
Black, dotted: Forecasts based on AR(1) models for the differenced yields

Gray, solid: Random-walk forecasts based on fitted 3-factor model

Gray, dashed: Forecasts based on AR(1) models for the factors

Gray, dotted: Forecasts based on AR(1) models for the differenced factors

Figure 5



Notes: Cumulative absolute errors of 12-months-ahead forecasts of monthly yields at different maturities relative to random-walk benchmark (a: 3 months, b: 12, c: 36, d: 120, e: average over all 17 maturities; f: average over all 17 maturities for cumulative squared errors; black vertical line: start of forecasting period of Diebold and Li, 2006).

Black, solid: Random-walk forecasts based on yields

Black, dashed: Forecasts based on AR(1) models for the yields

Black, dotted: Forecasts based on AR(1) models for the differenced yields

Gray, solid: Random-walk forecasts based on fitted 3-factor model

Gray, dashed: Forecasts based on AR(1) models for the factors

Gray, dotted: Forecasts based on AR(1) models for the differenced factors

Our results suggest that factor models do, in general, not allow more accurate term-structure forecasting. The forecasts based on AR(1) models for the factors are particularly bad. The common practice of using them as benchmarks for the evaluation of new forecasting

methods should therefore be discouraged in order to avoid overly optimistic conclusions. An obvious alternative would be to use forecasts based on AR(1) models for the differenced yields (in addition to the random walk forecasts). Apart from suitable benchmarks, the length of the forecast period is also critical. Anything can happen in a short period of a few years. The more sophisticated a forecasting method is, the longer the forecast period should be. For example, when the factor models are fitted locally in order to ensure stationarity, sophisticated testing procedures are required, which typically depend on a set of hyperparameters (Chen and Niu, 2014). In such a case, a forecast period of about ten years might probably be too short, particularly when the comparison of the competing forecasts is based on a single numerical measure rather than on the whole time course.

Clearly, increasing the length of the time series just by switching from monthly to daily data does not necessarily help a great deal when the forecast horizon remains unchanged. However, we could still try to construct additional predictors from the daily data either in a well-thought-out manner (e.g., local sample moments such as variances, autocovariances and cross-covariances) or, more erratically, with the help of machine learning algorithms. A further issue for future study is to augment conventional measures of forecast accuracy such as the MSE and the MAE with a measure of economic usefulness (e.g., for the prediction of economic downturns since the term spread is a strikingly accurate predictor of future economic activity; see Bauer and Mertens, 2018; for further interactions between the macroeconomy and the yield curve see Diebold *et al.*, 2006).

Acknowledgements

The authors thank the managing editor and two anonymous reviewers for helpful suggestions and comments.

References

- Bauer, M.D. and Mertens, T.M., 2018. Economic Forecasts with the Yield Curve. *FRBSF Economic Letter*, 2018-07.
- Chen, Y. and Niu, L., 2014. Adaptive dynamic Nelson–Siegel term structure model with applications. *Journal of Econometrics*, 180, pp.98–115.
- Christensen, J.H.E, Diebold, F.X. and Rudebusch, G.D., 2011. The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics*, 164, pp. 4–20.
- Cochrane, J.H. and Piazzesi, M., 2005. Bond risk premia. *American Economic Review*, 94, pp.138-160.
- Diebold, F.X. and Li, C., 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130, pp. 337-364.
- Diebold, F.X. and Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, pp. 253-263.
- Diebold, F.X., Rudebusch, G.D. and Aruoba, B., 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131, pp. 309-338.
- Fama, E. and Bliss, R., 1987. The information in long-maturity forward rates. *American Economic Review*, 77, pp.680–692.
- Favero, C.A., Niu, L. and Sala, L., 2012. Term structure forecasting: no-arbitrage restriction vs. large information set. *Journal of Forecasting*, 31, pp. 124–156.

- Filipović, D., 1999. A note on the Nelson–Siegel family. *Mathematical Finance*, 9, pp. 349-359.
- Harvey, D., Leybourne, S. and Newbold, P., 1997. Testing the equality of prediction mean squared errors, *International Journal of Forecasting*, 13, pp. 281-291.
- Kaya, H., 2013. Forecasting the yield curve and the role of macroeconomic information in Turkey. *Economic Modelling*, 33, pp. 1–7.
- Litzenberger, R., Squassi, G. and Weir, N., 1995. Spline models of the term structure of interest rates and their applications. *Working Paper*, Goldman, Sachs and Company.
- Matsumura, M.S., Moreira, A.R.B. and Vicente, J.V.M., 2010. Forecasting the yield curve with linear factor models. *Working Papers Series 223*, Central Bank of Brazil, Research Department.
- Mvingi, J. and Takudzwa, K., 2014. Estimation of term structures using Nelson-Siegel and Nelson-Siegel-Svensson: a case of a Zimbabwean bank. *Journal of Applied Finance & Banking*, 4, pp.155-190.
- Nelson, C.R. and Siegel, A.F., 1987. Parsimonious modeling of yield curve. *Journal of Business*, 60, pp.473–489.
- Niu, L. and Zeng, G., 2012. The Discrete-Time Framework of the Arbitrage-Free Nelson-Siegel Class of Term Structure Models. Available at <http://dx.doi.org/10.2139/ssrn.2015858>
- R Core Team, 2017. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Sambasivan, R. and Das, S., 2017. A Statistical Machine Learning Approach to Yield Curve Forecasting. arXiv:1703.01536
- Shang, Y. and Zheng, T., 2018. Fitting and forecasting yield curves with a mixed-frequency affine model: Evidence from China. *Economic Modelling*, 68, pp. 145–154.
- Svensson, L.E.O., 1994. Estimating and interpreting forward interest rates: Sweden 1992-4, *CEPR Discussion Paper Series No 1051*.
- Svensson, L.E.O., 1995. Estimating forward interest rates with the extended Nelson and Siegel Method, *Sveriges Riksbank Quarterly Review*, 3, pp.13-26