Forecasting the yield curve: art or science?

Tomas K. Molenaars and Nick H. Reinerink and Marcus A. Hemminga

RiskCo BV, Utrecht, The Netherlands, RiskCo BV, Utrecht, The Netherlands, RiskCo BV, Utrecht, The Netherlands

12 March 2015

Online at https://mpra.ub.uni-muenchen.de/63526/
MPRA Paper No. 63526, posted 10 April 2015 14:48 UTC
FORECASTING THE YIELD CURVE: ART OR SCIENCE?

Abstract

The objective of our work is to analyze the forecast performance of the dynamic Nelson-Siegel yield curve model and, for comparison, the first order autoregressive (AR(1)) model applied to a set of US bond yield data that covers a large timespan from November 1972 to December 2008. As a reference we take the random walk model applied to the yield data. For our analysis, we make use of a simple parameter representing the relative forecast performance to compare forecasting results of different methods. Our findings indicate that none of the yield curve models convincingly beats the random walk model. Furthermore, our results show that deriving conclusions on basis of model testing for a limited time period is inadequate.

Introduction

A yield curve (i.e., the term structure of interest rates) represents the relationship between interest rates and the remaining time to maturity. Forecasting of the yield curve will provide important information for monetary policy, as it is a basis for investment and saving strategies. In this view, the development of models for forecasting yield curves is of fundamental importance to banks and financial institutions, such as life insurers and pension funds.

For modeling the zero-coupon yield Diebold and Li (2006) constructed forecasting models based on the Nelson-Siegel model (Nelson and Siegel, 1987) and tested the forecast performance using US Treasury bond yields. This dynamic Nelson-Siegel model (Be Pooter, 2007; Christensen et al., 2009) utilizes a set of exponential components whose contributions are analyzed as a function of time. This method, in fact, is based on modeling the yield curve using the adjusted Nelson-Siegel model (1994). It was found that this approach forecasts well, especially for 6 and 12-month forecast horizon. This success has given rise to the popularity of the dynamic Nelson-Siegel model in forecasting studies of yield curves. However, the question is: how well does this model perform over a large time period?

Theory and methodology

The models that we use in the forecasting procedures are summarized in Table 1. In the case of the dynamic Nelson-Siegel model, the yield curve is fitted with the following equation:

\[ y(t) = \beta_1(t) + \beta_2(t)e^{-\lambda_1 t} + \beta_3(t)e^{-\lambda_2 t} - \beta_3(t)e^{-\lambda_3 t} \]

(1)

Here we have four time-dependent parameters, which can be interpreted as follows: the shape parameter \( \beta_1 \) governs the exponential decay rate and parameters \( \beta_2 \) and \( \beta_3 \) represent the contribution of the so-called long-term component, short-term component and medium-term component, respectively. Eq. (1) is not linear in \( \lambda_i \), hence for every time \( t \) we should estimate the parameters by a nonlinear fit. However, we follow the approach of Diebold and Li (2006), by fixing \( \lambda_i = k \). This avoids potentially challenging numerical optimizations. Doing this enables us to estimate the remaining parameters \( \beta_i \) by ordinary least-squares regression. The resulting time series for these parameters are modeled subsequently using the AR(1) model.

In the forecasting procedures with the dynamic Nelson-Siegel model in Eq. (1), the AR(1) forecast for the parameters \( \beta_i,k \) is given by:

\[ \hat{\beta}_{i,k,t+h} = \hat{\beta}_{i,k,t} + \rho \times \left( \beta_{i,k,t} - \hat{\beta}_{i,k,t} \right) \]

(2)

where \( \hat{\beta}_{i,k,t} \) and \( \hat{\beta}_{i,k,t} \) are the estimated parameters and \( h \) is the forecast horizon. Assuming a constant value for \( \lambda_i \) the yield forecast curve at time \( t+h \) is given by:

\[ y_{t+h} = \beta_1(t+h) + \beta_2(t+h)e^{-\lambda_1 t} + \beta_3(t+h)e^{-\lambda_2 t} - \beta_3(t+h)e^{-\lambda_3 t} \]

(3)

To evaluate the out-of-sample performance of a forecasting procedure, we calculate the root-mean-square-error (RMSE), given by:

\[ \text{RMSE}_{\text{model}}(h) = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_{t+h} - \hat{y}_{t+h})^2} \]

(4)

where \( y(t) \) is the forecasted yield of the model, \( \hat{y}(t) \) is the yield from the data, and \( N \) is the interval of times for which the forecasts are made. The smaller the RMSE, the better the forecast quality of the model.

To be able to systematically compare the quality of the number of forecasting results of the models, we ‘compress’ them in terms of a relative forecast performance parameter \( F \). This parameter is defined as the relative difference in forecast error of the model with respect to the RW model:

\[ F_{\text{model}} = \frac{\text{RMSE}_{\text{model}}(h) - \text{RMSE}_{\text{RW}}(h)}{\text{RMSE}_{\text{RW}}(h)} \]

(5)

where \( \text{RMSE}_{\text{RW}}(h) \) is the RMSE of the random walk model and \( \text{RMSE}_{\text{model}}(h) \) is the RMSE value at all maturities \( \tau \) of the random walk model and fitting model, respectively.

We take the random walk model as our benchmark, as it has the most simple no-change forecast, to provide a minimum standard on predictive accuracy for each model. Possible values of \( F \) denote a better forecast of the model as compared to the random walk model; negative values indicate a reduced performance. By definition, the relative forecast performance of the random walk model is 0.

Results and discussion

Our forecasting results are presented in Fig. 1A and B, which show the relative forecast performance of the models NS and AR, respectively, as a function of time at different forecast horizons \( h \). In this figure on the horizontal axis, the starting dates are shown for the various forecast periods. For example, 1994 (see arrow) reflects the forecast study carried out by Diebold & Li (2006). This point indicates a forecast period from January 1994 to December 2000 (from 1994:1 to 2000:12, i.e., 84 months).

The advantage of using \( F \) is that it enables us to easily compare the forecasting results of different models applied to a large yield data set. However, a disadvantage is that valuable information about the effect of different maturity values \( \tau \) is lost.

Nevertheless, Fig. 1 demonstrates that the relative forecast performance offers an excellent way to analyze the overall trends in the forecasts at different forecast horizons.

Since the forecasting result of the dynamic Nelson-Siegel model depends on the value of \( \lambda \), its effect on \( \text{RMSE}_{\text{NS}} \) is investigated for different values of \( \lambda \) for a forecast horizon of 6 months. This result is presented in Fig. 2. As can be seen, taking other values for \( \lambda \) does not make much difference, except for \( \lambda = 0.03 \), which delivers poor forecasts in most cases. Again, this is another
FORECASTING THE YIELD CURVE: ART OR SCIENCE?

Abstract

The objective of our work is to analyze the forecast performance of the dynamic Nelson-Siegel yield curve model and, for comparison, the first order autoregressive (AR(1)) model applied to a set of US bond yields data that covers a large timespan from November 1971 to December 2008. As a reference we take the random walk model applied to the yield data. For our analysis, we make use of a simple parameter representing the relative forecast performance to compare forecasting results of different methods. Our findings indicate that none of the yield curve models convincingly beats the random walk model. Furthermore, our results show that deriving conclusions on basis of model testing for a limited time period is inadequate.

Introduction

A yield curve (i.e., the term structure of interest rates) represents the relationship between interest rates and the remaining time to maturity. Forecasting of the yield curve will provide important information for monetary policy, as it is a basis for investment and saving strategies. In this view, the development of models for forecasting yield curves is of fundamental importance to banks and financial institutions, such as life insurers and pension funds.

For modeling the zero-coupon yield Diebold and Li (2006) constructed forecasting models based on the Nelson-Siegel model (Nelson and Siegel, 1987) and tested the forecast performance using US Treasuries bond yields. This dynamic Nelson-Siegel model (Diebold, 2007; Christensen et al., 2009) utilizes a set of exponential components whose contributions are analyzed as a function of time. This method, in fact, is based on modeling the yield curve using a smooth function. It was found that this approach forecasts well, especially for a 6 and 12-month forecast horizon. This success has given rise to the popularity of the dynamic Nelson-Siegel model in forecasting studies of yield curve. However, the question is: how well does this model perform over a large time period?

To tackle this problem, we use a simple parameter representing the relative forecast performance with respect to the random walk models to facilitate the interpretation of the forecasting quality. We systematically examine the dynamic Nelson-Siegel model and the AR(1) model using the US Treasuries bond yields for an extensive historic data set ranging from November 1971 to December 2008. This data set is provided by Robert Bliss and covers the period from November 1971 (1971:11) to December 2008 (2008:12) with maturities 5, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months.

Theory and methodology

The models that we use in the forecasting procedures are summarized in Table 1. In the case of the dynamic Nelson-Siegel model, the yield curve is fitted with the following equation:

\[ y_t = \beta_0 + \beta_1 \left( \frac{1-e^{-\lambda t}}{\lambda} \right) + \beta_2 \left( \frac{1-e^{-\lambda t}}{\lambda} - e^{-\lambda t} \right) \]  

(1)

Here we have four time-dependent parameters, which can be interpreted as follows: the shape parameter \( \lambda \) governs the exponential decay rate and parameters \( \beta_0, \beta_1, \beta_2 \) represent the contribution of the so-called long-term component, short-term component and medium-term component, respectively. Eq. (1) is not linear in \( \lambda \), hence for every time 1 we should estimate the parameters by a nonlinear fit. However, we follow the approach of Diebold and Li (2006), by fixing \( \lambda = \kappa \). This avoids potentially challenging numerical optimizations. Doing this enables us to estimate the remaining parameters \( \beta_1 \) by ordinary least-squares regression. The resulting times series for the \( \beta_1 \) parameters are modeled subsequently using the AR(1) model.

In the forecasting procedures with the dynamic Nelson-Siegel model (Eq. (1), the AR(1) model for the parameters \( \beta_1 \), can be written as:

\[ \hat{\beta}_{t+h} = \hat{\alpha} + \hat{\beta} \beta_{t} \]  

(2)

where \( \hat{\alpha} \) and \( \hat{\beta} \) are the estimated parameters and \( h \) is the forecast horizon. Assuming the yield curve at time \( t+h \) is given by:

\[ y_{t+h} = \beta_0 + \beta_1 \left( \frac{1-e^{-\lambda (t+h)}}{\lambda} \right) + \beta_2 \left( \frac{1-e^{-\lambda (t+h)}}{\lambda} - e^{-\lambda (t+h)} \right) \]  

(3)

To evaluate the out-of-sample performance of a forecasting procedure, we calculate the root-mean-square-error (RMSE), given by

\[ \text{RMSE}_{\text{model}}(\tau) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_{t,i,h} - y_{t,i,h})^2} \]  

(4)

where \( \hat{y}_{t,i,h} \) is the forecasted yield of the model, \( y_{t,i,h} \) is the yield from the data, and \( t \) is the interval of times for which the forecasts are made. The smaller the RMSE, the better the forecast quality of the model.

To be able to systematically compare the quality of the huge number of forecasting results of the models, we "compress" them in terms of a relative forecast performance parameter \( F \). This parameter is defined as the relative difference in forecast error of the model with respect to the RW model:

\[ F_{\text{model}}(\tau) = \frac{\text{RMSE}_{\text{model}}(\tau) - \text{RMSE}_{\text{RW}}(\tau)}{\text{RMSE}_{\text{RW}}(\tau)} \]  

(5)

where \( \text{RMSE}_{\text{RW}}(\tau) \) is the RMSE value for forecast horizon \( \tau \) of the random walk model and fitting model, respectively.

We take the random walk model as our benchmark, as it has the most simple no-change forecast, to provide a minimum standard on predictive accuracy for each model. Positive values of \( F \) denote a better forecast of the model as compared to the random walk model; negative values indicate a reduced performance. By definition, the relative forecast performance of the random walk model is 0.

Results and discussion

Our forecasting results are presented in Fig. 1 and Table 1. We plot the relative forecast performance of the models NS and AR, respectively, as a function of time at different forecast horizons \( h \). In this figure on the horizontal axis, the starting dates are shown for the various forecast periods. For example, 1994 (see arrow) reflects the forecast study carried out by Diebold and Li (2006). This point indicates a forecast period from January 1994 which is four times December 2000 (from 1994:1 to 2000:12, i.e., 84 months).

The advantage of using \( F \) is that it allows us to easily compare the forecasting results of different models applied to a large yield data set. However, a disadvantage is that valuable information about the effect of different maturity values \( \tau \) is lost. Nevertheless, Fig. 1 demonstrates that the relative forecast performance offers an excellent way to analyze the overall trends in the forecasts at different forecast horizons.

Since the forecasting result of the dynamic Nelson-Siegel model depends on the value of \( \lambda \), its effect on \( F_{\text{NS}} \) is investigated for different values of \( \lambda \) for a forecast horizon of 6 months. This result is presented in Fig. 2. As can be seen, taking other values for \( \lambda \) does not make much difference, except for \( \lambda = 0.03 \), which delivers poor forecasts in most cases. Again, this another

---

By Tessa S. Heemstra, Wink H. Haenrikz and Marcus A. Hemmingsa

---

Table 1 Models used in the forecasting procedures.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Winter Walk model on the yield data</th>
<th>AR(1) model on the yield data</th>
<th>Dynamic Nelson-Siegel model, Eq. (1) and AR(1) on the ( \beta )-parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>Random walk model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td></td>
<td>AR(1) model</td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>Dynamic Nelson-Siegel model, Eq. (1) and AR(1) on the ( \beta )-parameters</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1: Relative forecast performance \( F \) of the models NS (\( F_{\text{NS}} \)) and AR (\( F_{\text{AR}} \)) (see Table 1) for forecast horizons 1, 3, 6, 9, and 12 months. Parameter \( \lambda \) is fixed at the value of 0.0669. The arrow at the year 1999 reflects the results of the forecast study carried out by Diebold and Li (2006).

Fig. 2: Effect of \( \lambda \) on the relative forecast performance \( F \) of the NS model for a forecast horizon of 6 months.
demonstration of the usefulness of the relative forecast performance parameter.

From the results shown in Fig. 1, a couple of interesting observations can be made.

In comparing $F_{NS}$ and $F_{AR}$ in Fig. 1, it can be seen that only for about 20% of the monthly data points between 1982 and 2002 $F_{NS}$ performs better than $F_{AR}$ (see the periods 1993-1995 and 2000-2002). This suggests that there is no convincing advantage in using the more advanced and complicated dynamic Nelson-Siegel model over a simple AR(1) model. This can be understood, because there are a couple of inherent weaknesses in using the dynamic Nelson-Siegel model.

Firstly, one can argue that the Nelson-Siegel curve (Eq. (1)) does not properly fit the yield curve at all dates (for a fixed value of $\lambda$). In fact, the Nelson-Siegel model imposes a functional form to the yield curve. If the yield curve does not fit to this form, the Nelson-Siegel model will result in inferior forecasts. It is well known that adding a fourth term to the Nelson-Siegel equation (the Svensson extension (Svensson, 1995)), which allows for a second "hump/trough", delivers a better yield curve fit. Although there is no fundamental economic theory that supports this Nelson-Siegel-Svensson equation, it is extensively used by Central Banks (BIS, 2005; Gilli et al., 2010). Conversely, in the four-term Nelson-Siegel-Svensson equation more parameters need to be fitted, increasing the risk of fitting noise arising from parameter correlation and multiple local optima (Hawkins, 2004; Gilli et al., 2010).

Secondly, in the estimation of the $\beta$-parameters, it is assumed that $\lambda$ is fixed. However, it is questionable whether the Nelson-Siegel equation with a fixed $\lambda$ will perform well in all cases. In Fig. 1, we have used a constant value of $\lambda$ of 0.0609 (in month$^{-1}$) that is optimized by Diebold and Li (2006) for the result at 1994. The findings in Fig. 2 reveal that the effect of varying $\lambda$ is small, thus the value of $\lambda$ will not affect the main conclusions obtained from Fig. 1. Even so, the assumption of a fixed $\lambda$ may be a source for the low overall relative forecast performance of the dynamic Nelson-Siegel model as compared to the forecast performance of the AR(1) model.

The most striking point in Fig. 1 is that for almost all monthly data points the relative forecast performance $F$ is negative, demonstrating that none of the models AR and NS can convincingly beat the random walk model. Thus the most simple random walk forecasting model performs the best.

Finally, our results clearly show that deriving conclusions on basis of model testing for a limited time period is inadequate.

Acknowledgments
We thank Dr. Robert Bliss (Wake Forest University, Winston-Salem, USA) for kindly providing the US yield data and Dr. Michel De Pooter (Federal Reserve Board of Governors, Washington, USA).

References

1 – In this paper it is argued that the value of $\lambda$ that maximizes the medium-term component in Eq. (2) at exactly 30 months is $\lambda = 0.0609$. This statement is incorrect. The medium-term component has a bump shape with a maximum at $\lambda = 1.759$. From this relationship, it can be seen that $\lambda = 0.0609$ actually corresponds to 29.44 months.

2 The most striking point in Fig. 1 is that for almost all monthly data points the relative forecast performance $F$ is negative, demonstrating that none of the models AR and NS can convincingly beat the random walk model. Thus the most simple random walk forecasting model performs the best.

3 Finally, our results clearly show that deriving conclusions on basis of model testing for a limited time period is inadequate.