ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrics and Operations Research

Modeling and Forecasting the Yield Curve Using Time-varying Parameters

J.J. de Vlieger (400967)

Supervised by:
Prof. D.J.C. van Dijk
Second assessor:
X. Ling

July 1, 2017

Abstract

We use the Nelson-Siegel framework to fit and model the dynamics of the yield curve. We compare models that allow for a time-varying decay parameter to the widely used two-step-approach that was introduced by Diebold and Li (2006). Next to the different models to fit the yield curve, we introduce a novice method, recently used in forecasting realized volatility, to model the latent factors over time. This method includes the standard errors of the factor coefficients that are obtained in the first step of the process. We find that there is a clear trade-off between the in-sample fit and the out-of-sample forecasting power. Robust factors prove to be essential for the out-of-sample forecasting power.

Contents

1 Introduction

The term structure of the yield curve has been of interest for market practitioners and academia for quite some decades. Accurately fitting and forecasting the term structure is of importance for bond portfolio managers as well as policy makers and risk managers. Various models have gotten a lot of attention over the past years. The models used to describe the yield curves can be roughly divided into three classes. The affine equilibrium models, see for example Vasicek (1977), and the no-arbitrage models, see for example Hull and White (1990), can be considered two popular methods to fit and forecast the yield curve. However, Duffee (2002) showed that these models do not outperform the Random Walk in an out-of-sample setting.

More recently, the contribution of Diebold and Li (2006) (hereafter, DL) gained a lot of attention. They use a variation of the parsimonious factor model as described in Nelson and Siegel (1987) (hereafter, NS), to obtain three latent factors that they interpreted as the level, slope, and curvature of the yield curve. These latent factors are then modeled overtime by means of (vector) autoregressive models. This approach is often referred to as a two-step approach. Where, in the first step one fits the yield curve to obtain latent factors. These factors are obtained by applying a simple regression, with the yields of various maturities as depended variable, for each point in time. The coefficients that are obtained for each independent variable in the regression can than be seen as time series. Then, in the second step, one can model these time-series over time and reconstruct the yield curve. This approach allows the user to both fit and model the yield curve over time and touch upon the yield curve dynamics. Diebold and Li found that their model significantly outperformed most of the benchmark models on the medium and longer forecast horizons. Since the work of DL, the research regarding modeling the yield curve has extended in various ways.

First of all, the two-step approach in DL was extended by Diebold, Rudebusch, and Aruoba (2006) (hereafter, DRA) by means of a state-space model. A state-space model allows the user to instantaneously model and interpret the three factors. After the work of DRA various researchers have used the state-space model to model the yield curve. Diebold, Li, and Yue (2008) use the state-space model to incorporate a global yield model, that models the yield curve for various countries. They find that global factors exist based on four developed countries, but that there are some differences between countries. Morita and Bueno (2008) used the same framework to look at emerging countries and find that there are linkages between developed and emerging countries. Moreover, Bae and Kim (2011) investigate the global and regional factors for Asian countries. Their results are in line with those of Diebold et al. (2008) and Morita and Bueno (2008) yet, very limited regional differences are noticed. The papers mentioned in this paragraph all focus on regional differences. Yu and Zivot (2011) however looks at US corporate bonds with different investment ratings including both investment grade and high yield bonds. They also us the State-space model, but find that only for high yield bonds with short maturities, the state-space models outperforms the two-stage model of DL in an out-of-sample setting. However, Caldeira, Moura, and Portugal (2010) finds that the state-space model outperforms the two-stage model of DL, when looking at data on Brazilian government bonds. Furthermore, Yu and Zivot find suggestions that parameters might be time-varying. Koopman, Mallee, and van der Wel (2010) use this idea and implement a state-space model that allows for time-varying parameters. They find that the in-sample fit is improved. However, we do not know whether this model forecasts better, since Koopman et al. do not focus on out-of-sample analysis.

Secondly, besides the state-space models, Yu and Salyards (2009) look at US corporate bonds with different investment ratings, using the two-step approach. They find that depending on the investment rating, different values of λ will give a better fit. They also find that λ might be time-varying, but that a good in-sample fit does not guarantee a good out-of-sample forecast. Almeida, Gomes, Leite, and Vicente (2007) also suggest that the value for the decay parameter in DL might not be optimal in each situation. They look at Brazilian government bonds and choose the value of λ based on one-step-ahead forecasts for a 'train' sample. Furthermore, they find that implementing a second curvature factor in the NS framework significantly outperforms the DL model.

A lot of research has been done on parsimonious factor models. However, to model and forecast the dynamics of the yield curve, most research assumes that the decay parameter λ is stable over time. Some research has been dedicated to time-varying parameters such as Koopman et al.. Also, Yu and Zivot find suggestions that λ might vary over time. An advantage of using a fixed value of λ is that it ensures that the latent factors are more stable over time. However, little is known about the effect of loosening the restrictions of this parameter. Instability in the latent factors might result in weak forecasting results, when applying autoregressive models. Bollerslev, Patton, and Quaedvlieg (2016), recently used a new model to forecast realized volatility in option prices that takes into account the error of estimation of the previous step. They proved that these types of models significantly outperformed the autoregressive models. Since this method is very recent and only tested for realized volatility, we are interested if this might also work for yield curve forecasting.

We will extend existing literature, regarding modeling the yield curve, in two ways. The first extension focuses on the first step of the DL framework, where we fit the yield curve. We will look at the dynamics at play when fitting the yield curve. We relax the assumption of DL and compare different restrictions on the Nelson-Siegel formula when fitting the yield curve. The second extension is related to modeling the latent factors and uses the idea of Bollerslev et al. (2016) to take into account the estimation error that we made in the first step. This paper is therefore an extension of the original paper by DL. Although the use of state-space models is very popular among academia, there is little evidence that state-space models outperform the original models of DL. Therefore we choose to only investigate the effect of time-varying parameters in the two-step approach.

The rest of this paper is structured as follows. Section 2 discusses the methods and data used to fit and model the yield curve. We start of by briefly introducing the Nelson-Siegel frame work, after which we discuss the data that we use in our further analysis. The section continues by describing how we intend to fit the yield curve and finally how we model the latent factors over time. In section 3 we analyze how each of the models, discussed in the previous section, is able to fit the yield curve. Furthermore, we analyze how the latent factors behave over time. Then, section 4 analyses how well each of the models performs in an out-of-sample setting. We will compare the different models that were used to fit the yield curve, as well as different models that were used to model the latent factors. Finally, in section 5 we conclude and discuss further research opportunities.

2 Methods and Data

In this section we describe the data that we use in our analysis as well as the methods used in order to fit the yields and model the latent factors that we obtain. We start of by introducing the Nelson-Siegel framework, which is the basis of fitting the yield curve. Next, we describe the data that we use in our analysis. The main part of this section discusses the way how the yield curve is fitted. The assumption of a stable decay parameter is compared to a method that allows for a time-varying λ. Also, we will introduce four other models that restrict the latent factors. This section ends by describing how we model the latent factors over time.

2.1 The Nelson-Siegel yield curve

The original formula introduced by NS in order to fit the yield curve is the following:

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

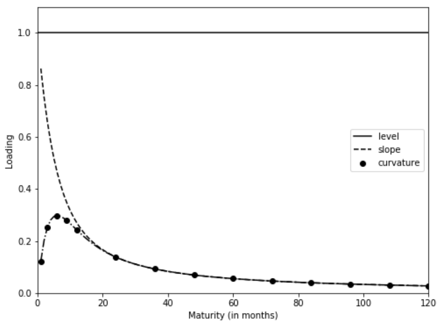
Where y_t denotes the continuously compounded zero-coupon nominal yield to maturity, β_{it} the three latent factors, which are widely interpreted as the level, slope and curvature of the yield curve, λ_t the decay parameter, and τ as the time until the bond matures, given in months. Throughout this paper, when we are talking about maturities we will always refer to the time to maturities denoted in months.

DL altered the original equation because the loadings of b_2 and b_3 have a similar monotonically decreasing shape. DL argue that this may result in multicolinearity and loss of interpretation of the factors. Therefore, DL use the following formula where $b_{1t} = \beta_{1t}$, $b_{2t} = \beta_{2t} + \beta_{3t}$, and $b_{3t} = \beta_{3t}$:

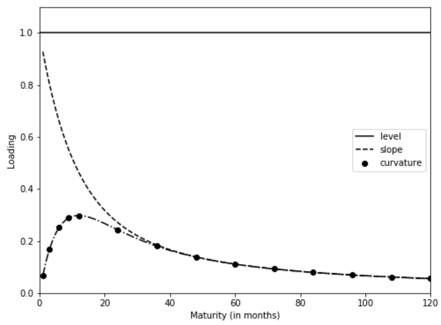
$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}\right) + \beta_{3,t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}\right) \tag{2}$$

As mentioned earlier, the parameters β_{it} represent the level, slope and curvature of the yield curve. Furthermore, DL show that they can also be interpreted as long-term, short-term and medium-term parameters, respectively. They also link the latent factors to the actual yields where $\beta_{1t} \sim y_t(120)$, $\beta_{2t} \sim y_t(120) - y_t(3)$, and $\beta_{3t} \sim 2y_t(24) - (y_t(120) + y_t(3))$, which we will refer to as the theorethical level, slope and curvature. The decay parameter, λ_t, in equations (1) and (2) influences the decay rate of the slope and curvature parameter. Small values for λ result in a slow decay, while large values imply a fast decay in the slope and curvature parameter. Furthermore, the value of the decay parameter controls where the curvature parameter reaches its maximum. In the work of DL the value of λ_t is assumed stable over time and has a value of 0.0609, which implies that the curvature parameter reaches its maximum at a maturity of thirty months.

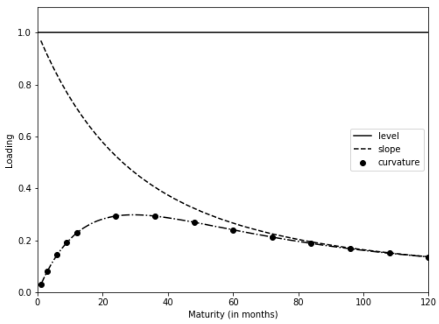
To visualize the effect of the decay parameter on the other parameters, figure 1 shows the factor loadings for different values of λ, plotted against the time to maturity. Here figure 1c shows the factor loadings as used by DL. The other three panels show the three factors where the value of λ is chosen such that the curvature parameter is maximized at 6, 12, and 60 months. Only the slope parameter and curvature parameter are affected by λ, but this is evident from equations (1) and (2).



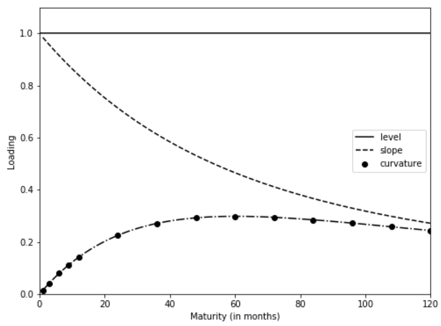
(a) $\lambda = 0.2989$, β_{3t} maximized at 6 months



(b) $\lambda = 0.1494$, β_{3t} maximized at 12 months



(c) $\lambda = 0.0609$, β_{3t} maximized at 30 months



(d) $\lambda = 0.0298$, β_{3t} maximized at 60 months

Figure 1: The three factor loadings (level, slope, and curvature) plotted against the time to maturity, given a certain value for λ. In each panel, the level is shown in the solid line, the slope in the dashed line, and the curvature in the dotted line. The level is equal to 1 for all maturities, the slope equals $\left(\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau}\right)$, and the curvature equals $\left(\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}\right)$. Where τ is the time to maturity in months.

2.2 Data

In our research, we use so-called "unsmoothed Fama-Bliss yields" for modelling and forecasting the yield curve. This name is derived from the researchers (Fama and Bliss (1987)) who came up with this, now very popular, approach to constructing yield curves. The "unsmoothed Fama-Bliss yields" are obtained by averaging the corresponding "unsmoothed Fama-Bliss forward rates". These rates are obtained from bonds that are taxable and non-calable, measured on a monthly basis. For the complete methodology on constructing "unsmoothed Fama-Bliss yields" we refer to the original work of Fama and Bliss. Throughout this paper we will refer to the "unsmoothed Fama-Bliss yields", as yields.

For our research, we gratefully use the dataset that was provided by van Dijk, Koopman, van der Wel, and Wright (2014). This data set considers the "unsmoothed Fama-Bliss yields" for the U.S treasury given on a monthly basis. Basically, it is an extension of the database that was used originally by DL. It is extended in the sense that the data starts earlier, also covering a period before 1985, and ends later, in the end of the Financial economic crises from 2007 and 2008. The complete sample runs from January 1970 (1970:M01) until December 2009 (2009:M12) and includes the following (fixed) maturities 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months.

Table 1 shows the descriptive statistics of the data as well as a theoretic level, slope, and curvature that were introduced above. Where the theoretic level is equal to the yield on the ten years bonds, the slope equals the difference between the ten years and three months bond yield, and the curvature as the difference between two times the two years bond yield and the sum of the three months and ten years bond yield. From table 1 we can clearly derive that the yields are on average upward sloping and become more persistent for longer maturities. This is also reflected in the autocorrelations. Furthermore, we observe that the slope and curvature show a relatively high standard deviation compared to their sample mean and the theoretical level.

Table 1: Descriptive statistics, yields

Maturity (Months)	Mean	Std. dev.	Min.	Max.	$\hat{\rho}1$	$\hat{\rho}12$	$\hat{\rho}30$
3	5.766	3.071	0.041	16.019	0.983	0.785	0.448
6	5.969	3.098	0.150	16.481	0.984	0.800	0.482
9	6.083	3.089	0.193	16.394	0.985	0.809	0.509
12	6.166	3.053	0.245	16.101	0.985	0.815	0.528
15	6.253	3.029	0.377	16.055	0.985	0.824	0.553
18	6.324	3.009	0.438	16.219	0.986	0.832	0.574
21	6.387	2.990	0.532	16.173	0.986	0.836	0.590
24	6.418	2.943	0.532	15.814	0.986	0.839	0.605
30	6.512	2.878	0.819	15.429	0.986	0.848	0.629
36	6.600	2.832	0.978	15.538	0.987	0.853	0.646
48	6.756	2.755	1.019	15.599	0.988	0.860	0.675
60	6.852	2.671	1.556	15.129	0.988	0.867	0.697
72	6.964	2.638	1.525	15.108	0.989	0.875	0.714
84	7.026	2.573	2.179	15.024	0.989	0.871	0.723
96	7.069	2.536	2.105	15.052	0.990	0.880	0.728
108	7.095	2.519	2.152	15.114	0.990	0.881	0.730
120 (level)	7.067	2.465	2.679	15.194	0.989	0.869	0.723
slope	1.301	1.362	-3.191	3.954	0.939	0.432	-0.130
curvature	0.003	0.863	-2.174	2.905	0.880	0.467	0.147

Note We show the mean, standard deviation, minimum, maximum, and sample autocorrelations of the "unsmoothed Fame-Bliss yiels" for all available maturities, given in months. The descriptive statistics are calculated using the full sample from 1970:M01 until 2009:M12. Level, slope, and curvature relate to the theoretical level, slope, and curvature. Where the theoretical level is equal to $y_t(120)$, the theoretical slope is equal to $y_t(120) - y_t(3)$, and the theoretical curvature is equal to $2y_t(24) - (y_t(120) + y_t(3))$.

2.3 Fitting yield curves

To fit the yield curve we apply least squares to equation (2). We apply this regression for each point in time, where we use the set of all available maturities. This means that for every month we have 17 observations of yields and their corresponding time to maturity. Since the main goal of this paper is to discover the effect of relaxing λ, we compare a model with a fixed λ to a model with a time-varying λ.

For the model where we fix λ we use the same value as proposed in DL, namely $\lambda_t = 0.0609$ for all t. As a result of fixing λ equation (2) becomes linear in β_{it}. Therefore, we can apply ordinary least squares (OLS) to obtain the coefficient values. For each day we obtain three coefficient values that we stack in time series. Later on we will model these time series, in order to come up with forecasts for the yields. A model that allows for time-varying parameters is somewhat more complicated, since relaxing λ results in that equation (2) becomes nonlinear in the parameters. Therefore we will apply nonlinear least squares (NLS) to fit the yield curve. This comes down to minimizing the following function for each point in time:

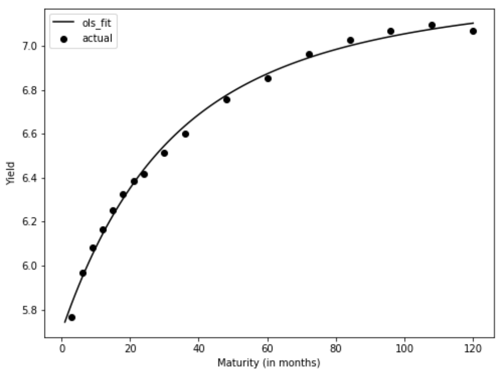
$$\min S_t(\beta) = \sum\nolimits_{\tau \in m} (y_t(\tau) - f(\tau, \beta_t, \lambda_t))^2 \tag{3}$$

With $y_t(\tau)$ being the observed "unsmoothed Fama-Bliss yields", τ the maturity in months, with $\tau \in m = \{3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 120\}$, and where:

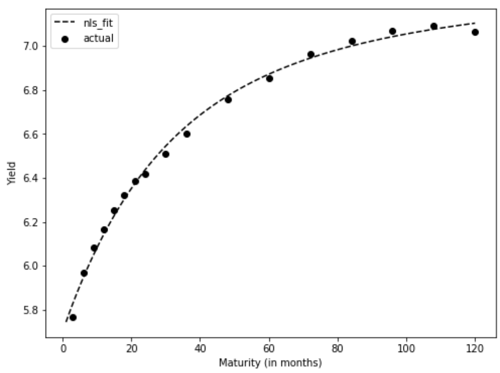
$$f(\tau, \beta_t, \lambda_t) = \beta_{1,t} + \beta_{2,t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}\right) + \beta_{3,t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}\right) \tag{4}$$

Equation 3 is optimized by using the Levenberg–Marquardt algorithm. This results in four parameter $(\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\lambda}_t)$ values that we obtain for every month, such that we obtain four time series. We choose the parameter values obtained from the OLS regression to be the starting values for the NLS regression.

To verify that on average these models both capture the yield curve, figure 2 shows the fit of the average yields for each maturity, that are given in table 1. Panel 2a shows the fit for the yield curve by using OLS and a value of λ equal to 0.0609 Panel 2b shows the fit using NLS. From figure 2 it is clear that both models give approximately the same fit. But, more interestingly is whether there is a difference in the parameter values. For the OLS regression these are 7.350, -1.651, -0.152, 0.0609 for β_{1t}, β_{2t}, β_{3t}, and λ_t respectively. For the NLS regression the parameter values to fit the yield curve are 7.354, -1.650, -0.0003, and 0.0551 respectively. The level parameter and slope parameter are very close to each other, however we observe a larger difference between the curvature parameter and decay parameter. Before we conclude anything about the two models, let us first explore the dynamics of the parameters in more detail.



(a) OLS



(b) NLS

Figure 2: Actual (data-based) yields, given in table 1 and the fitted (model-based) average yields. The x-axis shows the time to maturity in months, whereas the y-axis shows the yield.

Plotting the parameters, obtained from the OLS and NLS model, over time shows that the two models behave differently from time to time. In figure 3 we see that for certain periods the factor values are very different from each other.

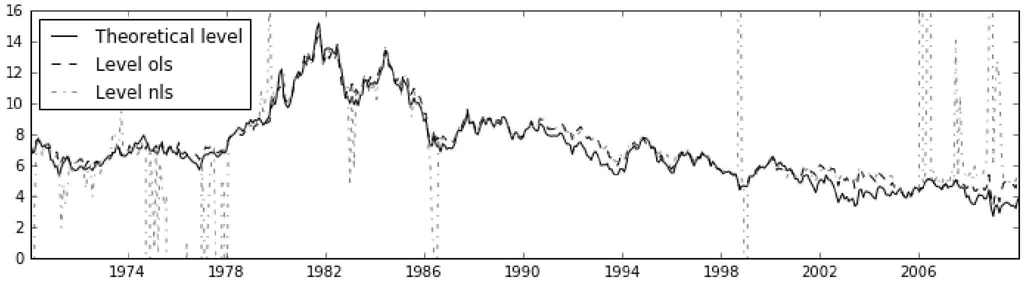
Panel (a) shows the level factor of both the OLS and the NLS regression, together with the theoretical level. For most observations the OLS and NLS approach have a similar course. However, we see some extreme observations in certain periods. The two periods with the most extreme observations are the periods from 1974 until 1978 and from 2006 until 2009. The correlation between the theoretical level and the level parameter obtained from the OLS and NLS model are 0.98 and 0.04 respectively. The correlation between the OLS and NLS model for the level parameter is even lower, 0.03. These low correlations are most likely due to the extreme values that we observe in the parameter values of the NLS regression. Panel (b) shows the time series of the inverse of the slope parameter, $-\beta_{2t}$, obtained from the OLS and NLS model. The inverse is taken because we defined the difference of the theoretical slope as $y_t(120) - y_t(3)$. The correlations of the theoretical slope and β_{2t} are -0.98 and -0.05, for the OLS and NLS method respectively. The correlations between the OLS and NLS model 0.01. Panel (c) shows the curvature factor. Correlations with the theoretical curvature 0.99 and 0.07, for the OLS and NLS model respectively, whereas the correlations between the two models is 0.05. The final panel (d) shows the decay parameter λ of the OLS and NLS model. For the OLS model this value is fixed at 0.0609 over the full sample. The two decay factors have 0 correlation.

In contrary to what figure 2 might suggest, the factors obtained from the OLS and NLS model are very different for some time periods. Also when comparing the factors based on their correlation, we conclude that the two models are not alike. However, this does not say anything about how the two model are able to fit the yield curve. In fact, we would expect that the NLS model will fit the data better. Yet, the downside of the extreme values in the factors of the NLS model is that they are harder to predict and might lead to poor forecasting results. Another problem the NLS model provides, is that the factors are more correlated to each other, compared to the NLS model. For the NLS model the highest correlation is 0.38, whereas for the NLS model the highest correlation is -0.89.

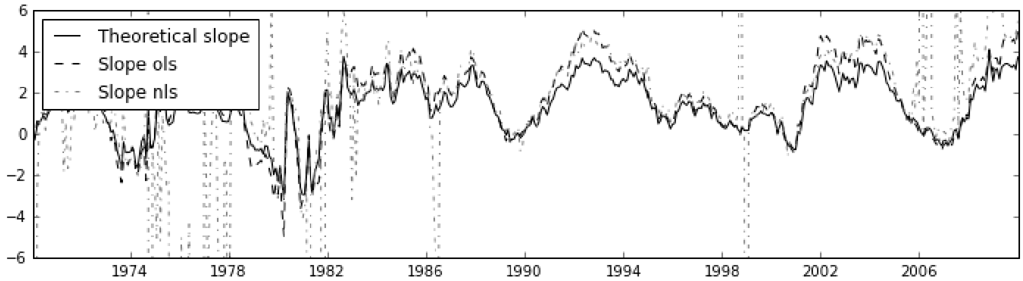
Because of the properties of the factors obtained from applying NLS, without imposing any restrictions, we will impose constraints on the coefficients. In the next section we will propose four new models that make a restriction on each of the factors in the NLS model. In section B we will look into the in-sample properties of all the models in more detail.

2.4 Restrictions on NLS

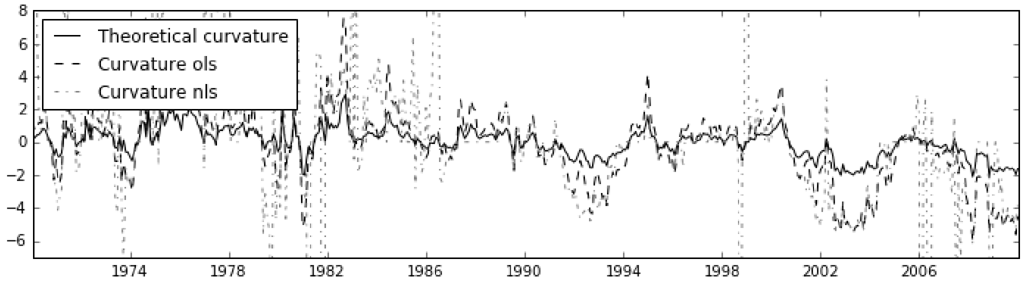
Figure 3 suggests that applying NLS without imposing restrictions leads to non-persistent factors. This might give a better in-sample fit, but it will also make forecasting more difficult. The question arises what kind of restrictions should be implemented in order to obtain more persistent factors that are easier to forecast, but on the other hand still give a good enough in-sample fit. In this section we will propose four new models. In all of the models we start from the unrestricted NLS model and impose a constraint on one of the four factors. In this way we are able to investigate the robustness of each factor on the fit of the yield curve as well as the effect on the forecasting power.



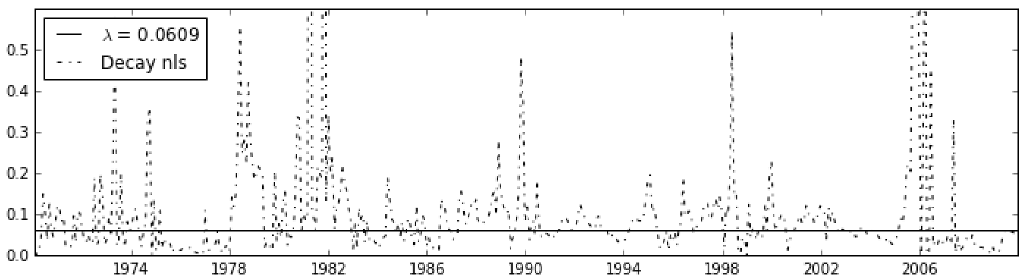
(a) Level



(b) Slope



(c) Curvature



(d) Decay (λ)

Figure 3: Parameter values for the level, slope, curvature, and decay. We show the parameter values obtained from the ordinary least squares (OLS) and the nonlinear least squares (NLS) model, as well as the theoretical level, slope and curvature. The theoretical level, slope, and curvarture are displayed with a black solid line. The OLS and NLS factors are displayed with a black and grey dashed line respectively. Panel (a) shows the level factor, β_{1t}, from the OLS and NLS method as well as the theoretical level. Panel (b) shows the inverse of the slope factor, $-\beta_{2t}$, as well as the theoretical slope. Panel (c) shows the curvature factor, β_{13}, as well as the theoretical curvature. Finally, panel (d) shows the decay parameter, λ_t, obtained from the NLS model, as well as the fixed value ($\lambda_t = 0.0609$) used in the OLS model.

For sake of notation we will refer to the unrestricted NLS model as Model 0. In Model 1 we impose a restriction on the level parameter. Because of the high correlation between the OLS level factor and the theoretical level factor we will restrict the level factor to be equal to the theoretical level, such that we will apply NLS to the following equation:

$$y_t(\tau) = y_t(120) + \beta_{2,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \tag{5}$$

In Model 2 and Model 3 we restrict the slope and curvature parameter using the same logic as in Model 1. We set β_{2t} and β_{3t} equal to the inverse of the theoretical slope and the theoretical curvature respectively. The two models are shown in equation (6) and (7).

$$y_t(\tau) = \beta_{1,t} + [y_t(120) - y_t(3)]\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \tag{6}$$

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + [2y_t(24) - [y_t(120) - y_t(3)]]\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \tag{7}$$

Finally, in model 4 we make a restriction on the parameter λ_t. In contrast to the other three models we do not fix the parameter, because this would simply be the OLS model. Therefore, we allow λ_t to vary over time but within boundaries, such that the curvature parameter reaches it maximum value between 24 and 36 months. These two values are most commonly used in previous research. The model looks as follows:

$$\begin{aligned} y_t(\tau) &= \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3,t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \\ \text{s.t.} &\quad 0.0498134 \leq \lambda_t \leq 0.0747201 \end{aligned} \tag{8}$$

2.5 Modelings the latent factors

After fitting the yield curve, we want to model the 4 latent factors over time in order to forecast the whole yield curve for different time horizons. Besides the standard autoregressive models we introduce a new model that is inspired on the work of Bollerslev et al. (2016). They refer to this model as an ARQ model, which they use to model and forecast realized volatility on a dense interval. The ARQ model is an autoregressive model where the lagged variable is linearly dependent on some error term. We gratefully use this idea to define our ARQ(1) model as:

$$\beta_t = \phi_0 + (\phi_1 + \phi_{1Q}\sqrt{Q_{t-1}})\beta_{t-1} + \epsilon_t \tag{9}$$

With $\sqrt{Q_t}$ being the standard error of the parameter β in the first regression at time t, and $\beta_t \in (\beta_{1t}, \beta_{2t}, \beta_{3t}, \lambda_t)$. One can easily see that the ARQ(1) becomes an AR(1) model when Q approaches 0, and penalizes, so to say, observations with a large error. In this sense we are actually weighing the past observations and give a lower weight to dates where we were not able to get a perfect fit of the yield curve, and higher weights to observations where we fit the yield curve better. Therefore we expect ϕ_{1Q} to be negative.

A potential downside of this way of modeling the factors is that if the standard error becomes to big it might occur that $\phi_1 + \phi_{1Q}\sqrt{Q_t} < 0$. To solve this we can incorporate a logistic function to model the factors, like in equation (10). We refer to this model as ARQL(1).

$$\beta_t = \phi_0 + \frac{\phi_1}{1 + exp(\phi_{1Q}(\sqrt{Q_{t-1}} - c))}\beta_{t-1} + \mu_t \tag{10}$$

With c being a constant. We choose c to be the average of the standard errors of the three factors obtained from applying OLS.

We compare the ARQ(1) and the ARQL(1) model to a simple autoregressive model with one lag (AR(1)), which has proven to be a hard benchmark to beat in previous research, and a Random Walk (RW) model. We define the AR(1) and the RW model in equation (11) and (12) respectively.

$$\beta_t = \phi_0 + \phi_1\beta_{t-1} + \eta_t \tag{11}$$

$$\beta_t = \beta_{t-1} \tag{12}$$

3 In-sample Analysis

In this section we analyze the in-sample statistics of the models that we use to fit the yield curve in more detail. In total we analyze six models. The model where we obtained the factors by applying OLS, and five other models where we applied NLS (Model 0 up to and including Model 4). We analyze the models based on their descriptive statistics of the factors and later by analyzing the residuals of the fit.

3.1 Factor analysis

Table A5 (appendix A) shows the mean, standard deviation, minimum, maximum, autocorrelations, and the p-value of the Augmented Dickey-Fuller (ADF) test statistic. In panel (a) and (b) we find the summary statistics of the factors that were obtained from the OLS and (unrestricted) NLS regression. It is evident that the mean of all factors from the NLS regression are different from those obtained from the OLS regression. Also, we find that the decay parameter λ is much higher than the 0.0609 that was used in the OLS regression. Also, we see that the minimum and maximum values that the factors reach are much larger in the NLS model. This also explains the high standard deviation and the low autocorrelations. In fact, we find that the 12-month and 30-month lag autocorrelations are very close to zero.

It is clear that the NLS model will gives extreme values. In panels (c)-(f) we show the four models with a restriction on one of the factors. Panel (c) show the summary statistics for Model 1, where we set the level factor equal to the theoretical level. The mean of the slope and curvature factor seem somewhat more reliable, however, from the standard deviation as well as the minimum and maximum values we conclude that extreme values still exists. Also, the decay parameter is higher than in the unrestricted NLS model. Furthermore, the auto correlations for all lags are very close to zero. In Model 2 (panel (d)) we set the slope parameter equal to the theoretical slope. We now see that the mean for all factors are approaching those of the OLS model. Furthermore, we see that the level parameter as well as the decay parameter move between reasonable bounds. The curvature parameter still shows some large negative values. However, these are limited since the standard deviation is not even three times that of the OLS model. Although the second model seems to remove quite some extreme values, Model 3 does not succeed at this point. The level, slope, and decay parameter again show quite some extreme values, although less so as in the NLS model and Model 1. Finally, Model 4 shows again promising results. The mean of all four factors are closely related to that of the OLS model, there are no extreme values across the whole sample and we find that the autocorrelations are significantly different from zero for the first three factors. It is clear that restriction on the decay parameter λ works best in reducing extreme values. Although, it has to be noted that it remains still hard to say what the precise effect of each restriction is on the in-sample fit and out-of-sample prediction power.

The last column in table A5 shows the p-value of the ADF test. This test whether there is a unit root in the time series. This is an indication whether a time series is stationary or not. We conclude that most of the time series are stationary. Therefore, we will model the factors without taking the first difference.

3.2 Residual analysis

To evaluate how well the models fit the yield curve, table B6 (appendix B) shows the in-sample residual statistics. We show the mean, standard deviation , mean absolute error (MAE), root mean squared error (RMSE), and the autocorrelations of the errors on a yield level. To obtain these numbers we first calculated the yields by using the factors obtained from the regression to fit the yield curve. The errors are then calculated by taking the difference of the actual observed data points and the fitted yields for every given month. We choose to only display the 3, 12, 36, 60, and 120 month yields. Together they represent the short-, medium-, and long-term maturities, and therefore give a representative picture of the whole set of maturities.

Let us first compare the OLS and NLS models. The MAE and RMSE is lower for the NLS model for all maturities compared to the OLS model. Also, the errors seem to be less correlated for most maturities. Both models fit the medium-term maturities best and the short term maturities worst. Finally, the errors obtained from the NLS method are more persistent since they show a lower standard deviation. These results show that although the NLS method might give some extreme values in the factors, they do result in a better overall in-sample fit.

Now we want to see what the effect of the restrictions on the NLS model is on the fit of the yield curve. Model 1, where we fix the level parameter, shows the highest MAE and RMSE for the yields of a 120-month maturity bond. This is interesting, since we take the actual value of these bonds as an explanatory variable in the regression. One explanation might be that the model allows for low value for λ, however this might also occur in the unrestricted NLS. Also, we find the highest correlation in the errors for the 120-month maturity. In general, Model 1 gives a worse fit than the unrestricted NLS model. For the short-term maturities it still gives a better fit than the OLS model, however for the long-term maturities the OLS model performs better. Model 2 clearly under performs compared to the NLS model. Since the theoretical slope can be seen as the short-term factor, as argued by DL, we would expect the best fit for the short-term maturities. However, we see that the maturities that are used in the regression, in this case the 3-month and 120-month maturities, give the worst fit. Again we see relatively high autocorrelations in the errors for the short-, and long-term maturities. Model 3 shows the same pattern as Model 2. Again, we see the highest autocorrelations for the short-, and long-term maturities. The model clearly under performs to the unrestricted NLS model based on the RMSE. The RMSE is however, slightly lower compared to the OLS model except for the 3-months maturity. Finally, Model 4, scores better than the OLS model, but worse than the NLS model based on the RMSE.

Overall, the unrestricted NLS gives the best in-sample fit compared to all other models. Yet, we saw in table A5 that this gives very sensitive factors. Clearly there is a trade-off between in-sample fit and the sensitivity of the factors. In the next section we will see how this sensitivity influences the out-of-sample forecasting power.

4 Out-of-Sample Analysis

In this section we show how well the models, described in previous section, are able to capture the yield curve dynamics over time.

4.1 Forecasting methodology

For simplicity we we refer to the latent factors $(\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\lambda}_t)$ as $\hat{\beta}_t$. We model the latent factors $\hat{\beta}_t$ in three different ways. A simple AR(1) model, an ARQ(1) model, and an ARQL(1) model that were described in section 2.5. For the AR(1) model, forecasts for the latent factors are obtained as follows:

$$\hat{\beta}_{t+h|T} = \hat{\phi}_0 + \hat{\phi}_1 \beta_T \tag{13}$$

For the ARQ(1) model:

$$\hat{\beta}_{t+h|T} = \hat{\phi}_0 + (\hat{\phi}_1 \sqrt{Q_T}) \beta_T \tag{14}$$

And for the ARQL(1) model:

$$\hat{\beta}_{t+h|T} = \hat{\phi}_0 + \frac{\hat{\phi}_1}{1 + exp(\hat{\phi}_{1Q}(\sqrt{Q_T} - c))} \beta_T + \mu_t \tag{15}$$

Using the forecasted factor values we are able to reconstruct the yield curve. Forecasting errors are then obtained by comparing the forecasts to the actual yield rates. We then evaluate and compare the different models based on their RMSE. We define the RMSE as:

$$RMSE(\tau) = \sqrt{\frac{\sum_{t=1}^{T-h} (y_{t+h}(\tau) - \hat{y}_{t+h}(\tau))}{n}} \tag{16}$$

Where $\hat{y}_{t+h}(\tau)$ is defined as:

$$\hat{y}_{t+h}(\tau) = \hat{\beta}_{1,t+h} + \hat{\beta}_{2,t+h} \left(\frac{1 - e^{-\hat{\lambda}_{t+h}\tau}}{\hat{\lambda}_{t+h}\tau} \right) + \hat{\beta}_{3,t+h} \left(\frac{1 - e^{-\hat{\lambda}_{t+h}\tau}}{\hat{\lambda}_{t+h}\tau} - e^{-\hat{\lambda}_{t+h}\tau} \right) \tag{17}$$

We will look at three forecasting horizons: 1-month-ahead, 6-months-ahead, and 12-months-ahead. We use a moving window of ten years (120 months) to generate the forecasts. Furthermore, to keep all forecasting window equal we generate forecasts from 1981:M01 until 2009:M12, meaning that we have 348 forecasting points for each maturity. We will use a direct forecasting method for the 6-, and 12-months-ahead forecasting horizons.

4.2 Forecast evaluation

In tables 2, 3, and 4 we show the RMSE for all model combinations. Six models to fit the yield curve, and four models to model the factors. This results in a total of 24 models. For each model we show the RMSE for five maturities; 3, 12, 36, 60, and 120 months.

Table 2: 1-step-ahead forecast

	Maturity	OLS	NLS	Model 1	Model 2	Model 3	Model 4
Panel a: ARQL(1)							
	3	**0.479**	131.0	12.87	0.572	8.855	0.541
	12	**0.422**	119.5	6.607	0.518	6.194	0.525
	36	**0.400**	111.6	2.541	0.487	87.18	0.532
	60	**0.385**	111.3	1.539	0.429	3116	0.520
	120	**0.362**	112.1	0.807	0.366	8.6E+08	0.494
Panel b: ARQ(1)							
	3	0.476	47.91	8.402	0.559	9.089	**0.473**
	12	**0.415**	22.10	12.33	0.493	16.07	0.424
	36	**0.397**	7.790	153.7	0.449	23.95	0.401
	60	**0.381**	7.344	2009	0.404	26.18	0.384
	120	0.357	9.201	1.3E+06	**0.360**	27.87	0.357
Panel c: AR(1)							
	3	0.473	79.85	5.401	0.564	8.498	**0.469**
	12	**0.419**	8.872	1.284	0.499	19.52	0.426
	36	**0.399**	12.49	0.599	0.474	25.65	0.406
	60	**0.383**	18.19	0.452	0.426	26.92	0.390
	120	**0.359**	22.69	0.384	0.371	27.82	0.362
Panel d: Random Walk							
	3	0.465	**0.447**	0.458	0.488	0.455	0.461
	12	**0.414**	0.421	0.419	0.415	0.421	0.414
	36	0.404	0.394	0.412	0.407	**0.392**	0.402
	60	0.389	0.380	**0.378**	0.390	0.382	0.387
	120	0.365	0.369	0.427	0.368	0.379	**0.364**

Note We show the root mean squared error (RMSE) for all model combinations. Six models to fit the yield curve and obtain the latent factors, and four models to model the factors. This results in a total of 24 unique forecasts for the yields, that are obtained from a 1-month-ahead forecast. We show RMSE for five different maturities; 3, 12, 36, 60, and 120 months. The lowest RMSE for each maturity is depicted in bold.

The RMSE's for the 1-step-ahead forecasting results are shown in table 2. First let us compare the six models that were used to fit the yield curve. Overall, we conclude that the OLS model is superior to all other models. The Random Walk model however, does not give a clear winner, since the results are closely related. The results of the four models to model the factors depend on which model is used to obtain the factors. The Random Walk gives the lowest RMSE for the NLS model, Model 1, and Model 3. There is no clear winner among the other three models.

The 6-step-ahead forecast are shown in table 3. Again, we notice that the OLS model performs best for the ARQL(1) and AR(1) model. However, the forecasts from ARQ(1) model given the factors obtained from Model 4 shows a slightly lower RMSE for the short-, and long-term maturities. In general, we now see that the Random Walk model performs worse compared to the other three models, except when the factors from the NLS model or Model 3 are used. Model 1 performs better on the long-term maturities when using the ARQL(1) and AR(1) model. However, the ARQ(1) model performs very poorly, given the factors from Model 1.

Table 3: 6-step-ahead forecast

	Maturity	OLS	NLS	Model 1	Model 2	Model 3	Model 4
Panel a: ARQL(1)							
	3	**0.491**	130.9	13.25	0.595	7.931	0.604
	12	**0.435**	122.9	6.951	0.548	6.053	0.604
	36	**0.413**	116.8	2.658	0.513	1312	0.621
	60	**0.398**	116.2	1.608	0.451	1.6E+06	0.614
	120	**0.372**	116.2	0.835	0.384	1.4E+14	0.592
Panel b: ARQ(1)							
	3	0.502	158.9	7.495	8161	29.18	**0.496**
	12	**0.437**	26.38	47.72	1.4E+23	33.78	0.444
	36	**0.422**	220.4	1.4E+08	6.5E+73	38.50	0.425
	60	0.408	268.8	3.2E+16	3E+124	39.85	**0.407**
	120	0.376	221.9	8.7E+35	inf	40.90	**0.374**
Panel c: AR(1)							
	3	0.479	37.26	5.508	0.576	6.412	**0.476**
	12	**0.428**	9.439	1.448	0.513	14.05	0.437
	36	**0.406**	13.56	0.643	0.485	18.38	0.416
	60	**0.389**	19.88	0.470	0.437	19.26	0.399
	120	**0.364**	25.61	0.388	0.379	19.87	0.369
Panel d: Random Walk							
	3	1.250	**1.238**	1.240	1.259	1.243	1.248
	12	**1.183**	1.196	1.193	1.187	1.192	1.185
	36	1.081	1.069	1.077	1.081	**1.063**	1.079
	60	1.000	0.991	**0.988**	0.998	0.990	0.996
	120	0.896	0.905	0.923	**0.892**	0.917	0.897

Note We show the root mean squared error (RMSE) for all model combinations. Six models to fit the yield curve and obtain the latent factors, and four models to model the factors. This results in a total of 24 unique forecasts for the yields, that are obtained from a 6-month-ahead forecast. We show RMSE for five different maturities; 3, 12, 36, 60, and 120 months. The lowest RMSE for each maturity is depicted in bold.

Finally, the 12-step-ahead forecasts are shown in table 4. Again, we see that the OLS model performs best when modelling the factors with either the ARQL(1), ARQ(1), or the AR(1) model. Given the factors from the OLS model all these three models outperform the Random Walk model. The same three models also outperform the Random Walk model given the factors from Model 4. Model 2 also outperform the Random Walk model when using the ARQL(1) or AR(1) model. However, the ARQ(1) model provides poor forecasting results given the factors from Model 2. Model 1 and Model 3 perform poorly when using the ARQL(1), ARQ(1), or AR(1) model.

Table 4: 12-step-ahead forecast

Maturity	OLS	NLS	Model 1	Model 2	Model 3	Model 4
Panel a: ARQL(1)						
3	**0.504**	143.3	12.11	0.616	6.963	0.643
12	**0.448**	135.8	6.462	0.569	6.326	0.653
36	**0.423**	126.6	2.524	0.523	13.61	0.668
60	**0.405**	123.9	1.534	0.459	2346	0.661
120	**0.379**	122.1	0.799	0.392	4.3E+09	0.643
Panel b: ARQ(1)						
3	**0.512**	68.77	206.6	7.4E+04	53.13	0.519
12	**0.449**	112.5	2628	6.4E+26	49.54	0.465
36	**0.433**	287.3	6.8E+16	6.7E+84	41.78	0.441
60	**0.417**	307.0	6.5E+29	7E+142	35.45	0.420
120	**0.382**	227.9	1.4E+62	inf	24.15	0.382
Panel c: AR(1)						
3	0.483	28.70	5.109	0.590	2.361	**0.482**
12	**0.434**	10.23	1.376	0.528	5.456	0.445
36	**0.410**	14.24	0.653	0.493	8.038	0.423
60	**0.392**	20.46	0.475	0.442	8.678	0.403
120	**0.366**	26.19	0.385	0.383	9.281	0.371
Panel d: Random Walk						
3	1.981	**1.963**	1.963	1.975	1.975	1.979
12	**1.856**	1.869	1.868	1.864	1.862	1.857
36	1.611	1.601	1.604	1.608	**1.597**	1.610
60	1.458	1.451	1.451	1.457	**1.451**	1.456
120	1.316	1.322	1.333	**1.309**	1.333	1.316

Note We show the root mean squared error (RMSE) for all model combinations. Six models to fit the yield curve and obtain the latent factors, and four models to model the factors. This results in a total of 24 unique forecasts for the yields, that are obtained from a 12-month-ahead forecast. We show RMSE for five different maturities; 3, 12, 36, 60, and 120 months. The lowest RMSE for each maturity is depicted in bold.

Overall, we see that using an AR(1) given the factors from the OLS regression gives the best forecast for especially the longer forecasting horizons. We find that for the longer forecasting horizons the ARQL(1) and the ARQ(1) can obtain better forecasts than the Random Walk, however, in some situations they give very poor forecasting results. Model 1 and Model 3 do not give proper forecasting results. This might be due to the fact that we still obtain extreme values, although we restrict the NLS model. All these forecasts are obtained from the full forecast sample. In the next subsection we will compare two smaller sub-sample to see how these models perform at different time periods.

4.3 Sub-sample evaluation

So far we have only looked at the forecasting results over the full sample. The poor forecasting results might be due to the fact of extreme values. To see if this is actually the case we make two sub-samples. One that contains a lot of extreme values, and one that does not. The first sample, ranges from 1981M01-1985M12, and the second ranges from 2000M01:2004M12. Both samples contain 60 observations. Since we use a moving forecasting window of 120 months, we chose these two sample based on the amount of extreme values in the period before the two forecasting samples. The period before 1981 shows a lot of extreme values for the factors that are obtained from the models where we applied NLS. The period before 2000 however, does not have that many of these extreme values in the factors. Therefore, one would expect that the second sample gives better forecasting results.

Table C7 (Appendix C) shows the RMSE for these two sub-samples, for all three forecasting horizons. In the second sample the RMSE's are lower than in the first sample for all model combinations. However, the models that use the factors from the OLS regression still give the best forecasting results. Next, if we compare the ARQL(1), ARQ(1), AR(1), and Random Walk model we see that there is no clear winner among the models.

5 Conclusion

In this paper we attempt to fit the yield curve for any given point in time and attempt to model the yield curve over time. Our research extends the current literature in two ways. First of all, we discuss six models to fit the yield curve. Whilst different, all the models use the Nelson-Siegel formula as a starting point. The topic of interest is the decay parameter in this formula. Previous researchers chose to fix the decay parameter over time. We however, choose to allow for variance over time in the decay parameter λ. Next to that, we come up with four alterations to this model. In each of the four models we make a restriction on one of the parameters in the Nelson-Siegel formula. We find that all models are able to capture the yield curve. This does not come as a surprise since we relax the parameter values by allowing for a time-varying parameters. As a direct result that the parameters obtained from NLS are very sensitive and non robust. Setting the right constraints proves to be important when we want to forecast the yield curve.

Besides fitting the yield curve we are also interested in forecasting and modelling the parameters over time. We propose two new models, the ARQL and ARQ model. These two models take into account the standard error of the coefficients, obtained from the first regression. When comparing the out-of-sample forecasting results, by means of the

RMSE, we find that fitting the yield curve plays an important role. The model were OLS is applied to obtain the latent parameters proves to be a hard benchmark to beat. The results of the other models, where NLS is used, depend on the constraints. Simply applying NLS without any restrictions does not give a good forecast, this might be due to the fact that the factors can attain very extreme values over time. We find that some restrictions do give reasonable forecasts. Since we only use very basic restrictions on the parameters, we believe that further investigation in the relations between the parameters is of interest for further research.The novice ARQ(1) and ARQL(1) model that we use to model the yield curve dynamics over-time is very comparable to the AR(1) model, given that the parameters are well specified. As previously stated, these remain highly basics models. Because the parameters might be linked to each other, a more advanced version of the model that also takes into account the dynamics of the other parameters might improve the forecasting power of this method.

Overall, we conclude that there is a clear trade-off between the in-sample fit and the out-of-sample forecasting power. Perfectly fitting the yield curve does certainly not guarantee superior forecasting results. Over fitting the yield curve might in fact lead to very poor forecasting results. The AR(1) model given the factors obtained from an OLS regression with a fixed value of λ still gives the user good forecast results compared to the models we propose. However, it is interesting to investigate how these models perform when looking at daily, or even intra-day, data since the difference in the yield curve might then be smaller than when looking at monthly data. This might improve forecasting results, since the factors are expected to be more smooth over time. Furthermore, changing the forecasting window or using an indirect forecasting method might also lead to better out-of-sample results.

References

Almeida, C., Gomes, R., Leite, A., & Vicente, J. (2007). Does curvature enhance forecasting? *Working paper*.

Bae, B., & Kim, D. (2011). Global and regional yield curve dynamics and interactions: The case of some asian countries. *International Economic Journal*, *25*(4), 717–738.

Bollerslev, T., Patton, A., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*(1), 1-18.

Caldeira, J., Moura, G., & Portugal, M. (2010). Efficient yield curve estimation andforecasting in brazil. *Economia*, *11*(1).

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

Diebold, F., Li, C., & Yue, V. (2008). Global yield curve dynamics and interactions: A dynamic nelson–siegel approach. *Journal of Econometrics*, *146*(2), 351-363.

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

Duffee, G. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance*, *57*(1), 405-443.

Fama, E., & Bliss, R. (1987). The information in long-maturity forward rates. *The American Economic Review*, *77*(4), 680-692.

Hull, J., & White, A. (1990). Pricing interest-rate-derivative securities. *The Review of Financial Studies*, *3*(4), 573-592.

Koopman, S., Mallee, M., & van der Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, *28*(3), 329-343.

Morita, R., & Bueno, R. (2008). Investment grade countries yield curve dynamics. *Meeting of the Brazilian Econometric Society, December*.

Nelson, C., & Siegel, A. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, *60*(4), 473-489.

van Dijk, D., Koopman, S., van der Wel, M., & Wright, J. (2014). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*, *29*(5), 693-712.

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

Yu, W., & Salyards, D. (2009). Parsimonious modeling and forecasting of corporate yield curve. *Journal of Forecasting*, *28*(1), 73-88.

Yu, W., & Zivot, E. (2011). Forecasting the term structures of treasury and corporate yields: Dynamic nelson-siegel models evaluation. *International Journal of Forecasting*, *27*(2), 579-591.

Appendix A Factor summary

Table A5: Descriptive statistics, factors

	Mean	Std. dev.	Min.	Max.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
Panel a: OLS								
level	7.350	2.295	3.336	14.39	0.990	0.885	0.742	-1.189
slope	-1.651	1.843	-5.195	5.021	0.953	0.486	-0.140	-3.852*
curvature	-0.152	2.220	-6.032	7.523	0.888	0.456	0.098	-3.281**
Panel b: NLS								
level	-41.49	574.3	-8324	131.7	0.669	-0.004	-0.008	-7.642*
slope	64.74	651.4	-126.8	8329	0.515	-0.007	-0.010	-7.392*
curvature	32.93	667.2	-6502	8540	0.526	0.002	-0.003	-7.255*
decay	0.103	0.225	0.000	3.337	0.123	0.018	0.019	-5.831*
Panel c: Model 1, fixed $\hat{\beta}_{1t}$								
level	7.067	2.465	2.679	15.19	0.989	0.869	0.723	-1.318
slope	-7.781	130.2	-2846	4.180	-0.003	-0.003	-0.002	-21.90*
curvature	6.241	130.2	-6.737	2844	-0.002	-0.003	-0.005	-21.88*
decay	0.136	0.240	0.015	3.248	0.157	-0.026	-0.017	-5.096*
Panel d: Model 2, fixed $\hat{\beta}_{2t}$								
level	7.145	2.341	3.064	14.78	0.987	0.873	0.715	-1.378
slope	-1.301	1.362	-3.954	3.191	0.939	0.432	-0.130	-3.957*
curvature	-0.499	5.795	-101.6	7.506	0.259	0.153	0.002	-6.305*
decay	0.084	0.067	0.000	1.092	0.328	0.049	0.029	-5.619*
Panel e: Model 3, fixed $\hat{\beta}_{3t}$								
level	10.68	31.76	-100.8	400.0	0.586	-0.001	-0.010	-6.103*
slope	-5.252	32.26	-398.8	108.9	0.592	0.016	0.067	-5.922*
curvature	0.003	0.863	-2.174	2.905	0.880	0.467	0.147	-2.978**
decay	0.151	0.719	-0.016	10.50	0.179	-0.012	-0.024	-6.770*
Panel f: Model 4, restricted λ_t								
level	7.360	2.283	3.619	14.516	0.989	0.882	0.735	-1.331
slope	-1.673	1.841	-5.442	4.874	0.949	0.488	-0.126	-3.713*
curvature	-0.138	2.304	-6.084	6.817	0.867	0.495	0.115	-2.783***
decay	0.063	0.012	0.050	0.075	0.558	0.007	-0.054	-5.371*

Note We show the mean, standard deviation, minimum, maximum, sample autocorrelations, and the p-value for the augmented Dickey-Fuller (ADF) test for every factor and each model. In panel a, the decay factor is not mentioned because $\lambda_t = 0.0609$ for every t. Starts indicated significance level: 1%***, 5%**, 10%*. ADF H_0: there is a unit root, versus H_a: there is no unit root.

Appendix B Residual summary

Table B6: Descriptive statistics, residuals

Maturity	Mean	Std. dev.	Min.	Max.	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
Panel a: OLS									
3	-0.062	0.135	-0.794	0.295	0.103	0.149	0.733	0.357	-0.020
12	0.022	0.097	-0.224	0.475	0.073	0.100	0.604	0.340	-0.156
36	-0.036	0.065	-0.399	0.256	0.055	0.075	0.465	0.203	0.131
60	-0.021	0.085	-0.400	0.288	0.066	0.088	0.753	0.230	0.082
120	-0.036	0.138	-0.580	0.380	0.095	0.143	0.814	0.351	-0.052
Panel b: NLS									
3	-0.041	0.086	-0.547	0.256	0.060	0.096	0.614	0.203	0.158
12	0.011	0.075	-0.216	0.382	0.055	0.076	0.495	0.345	-0.023
36	-0.027	0.051	-0.193	0.249	0.044	0.058	0.332	0.155	0.089
60	-0.023	0.061	-0.267	0.242	0.048	0.065	0.548	-0.053	-0.079
120	-0.026	0.074	-0.443	0.245	0.054	0.078	0.451	0.056	-0.013
Panel c: Model 1, fixed β_{1t}									
3	-0.051	0.081	-0.561	0.315	0.068	0.096	0.547	0.160	0.022
12	0.022	0.078	-0.212	0.403	0.063	0.081	0.524	0.351	-0.049
36	-0.071	0.081	-0.324	0.252	0.084	0.108	0.649	0.406	-0.021
60	-0.012	0.088	-0.500	0.275	0.065	0.089	0.660	0.106	0.043
120	0.076	0.236	-0.559	0.807	0.182	0.247	0.895	0.653	0.175
Panel d: Model 2, fixed β_{2t}									
3	-0.126	0.129	-0.678	0.348	0.141	0.180	0.688	0.320	-0.157
12	0.036	0.075	-0.231	0.390	0.065	0.083	0.451	0.338	-0.132
36	-0.056	0.059	-0.382	0.274	0.067	0.082	0.425	0.220	0.006
60	-0.039	0.074	-0.276	0.290	0.068	0.084	0.731	0.247	0.158
120	0.011	0.142	-0.485	0.496	0.103	0.142	0.841	0.476	0.121
Panel e: Model 3, fixed β_{3t}									
3	-0.024	0.144	-0.653	0.399	0.100	0.146	0.742	0.367	0.025
12	0.005	0.094	-0.224	0.436	0.069	0.094	0.607	0.375	-0.154
36	-0.022	0.067	-0.292	0.267	0.052	0.071	0.540	0.143	-0.032
60	0.002	0.087	-0.389	0.356	0.064	0.087	0.758	0.274	-0.029
120	-0.068	0.153	-0.712	0.290	0.110	0.167	0.806	0.398	-0.155
Panel f: Model 4, restricted λ_t									
3	-0.056	0.120	-0.701	0.273	0.090	0.133	0.706	0.396	0.002
12	0.020	0.090	-0.217	0.437	0.067	0.092	0.573	0.348	-0.136
36	-0.036	0.063	-0.365	0.254	0.053	0.072	0.442	0.217	0.173
60	-0.019	0.074	-0.367	0.242	0.057	0.077	0.707	0.166	0.060
120	-0.037	0.123	-0.530	0.316	0.085	0.129	0.780	0.303	-0.072

Note We show the mean, standard deviation, minimum, maximum, mean average error (MAE), root mean squared error (RMSE), and sample autocorrelations for five different maturities for the six discussed models.

Appendix C Sub-sample forecast results

Table C7: Root mean squared errors for two sub-samples

1-month-ahead	OLS	NLS	Model 1	Model 2	Model 3	Model 4
Panel a: 1981:M01-1985M12						
ARQL(1)	**0,680**	230,0	0,847	0,750	1,106	0,831
ARQ(1)	**0,672**	40,56	0,780	0,712	1,738	0,679
AR(1)	**0,674**	58,32	0,814	0,756	3,886	0,676
RW	0,699	0,697	**0,696**	0,698	0,698	0,698
Panel b: 2000:M01-2004M12						
ARQL(1)	**0,308**	2,102	0,475	0,324	4,613	0,341
ARQ(1)	**0,303**	1,192	0,359	0,327	2,239	0,306
AR(1)	**0,304**	1,729	0,353	0,325	1,921	0,306
RW	0,299	**0,297**	0,338	0,307	0,303	0,298
6-months-ahead						
Panel a: 1981:M01-1985M12						
ARQL(1)	**0,697**	247,0	0,864	0,771	1,438	0,940
ARQ(1)	0,702	43,8	0,795	9,1E+21	2,084	**0,702**
AR(1)	**0,679**	37,9	0,832	0,758	4,142	0,683
RW	**1,764**	1,765	1,771	1,764	1,764	1,765
Panel b: 2000:M01-2004M12						
ARQL(1)	**0,314**	2,326	0,515	0,332	5,118	0,358
ARQ(1)	**0,309**	1,270	0,367	0,350	3,117	0,313
AR(1)	**0,311**	1,787	0,356	0,329	1,998	0,313
RW	0,845	0,845	0,855	0,848	**0,844**	0,845
12-months-ahead						
Panel a: 1981:M01-1985M12						
ARQL(1)	**0,719**	281,8	0,884	0,790	2,010	1,053
ARQ(1)	**0,722**	39,4	0,812	7,0E+13	2,666	0,726
AR(1)	**0,685**	35,5	0,854	0,770	4,411	0,691
RW	**2,651**	2,652	2,659	2,651	2,651	2,651
Panel b: 2000:M01-2004M12						
ARQL(1)	**0,313**	2,467	0,507	0,329	5,458	0,356
ARQ(1)	**0,308**	1,404	0,362	0,362	3,273	0,315
AR(1)	**0,313**	1,856	0,352	0,325	2,076	0,314
RW	1,372	1,372	1,373	1,374	**1,370**	1,372

Note We show the root mean squared error (RMSE) for all model combinations. Six models to fit the yield curve and obtain the latent factors, and four models to model the factors. This results in a total of 24 unique forecasts for the yields, that are obtained from a 1-, 6-, and 12-month-ahead forecasts. For each forecast horizon we compare two sub-samples. In panel (a) the sample ranges from 1981:M01-1985:M12, and in panel (b) the sample ranges from 2000:M01-2004:M12 We show RMSE of all maturities combines. The lowest RMSE for each maturity is depicted in bold.