UNIVERSITEIT ANTWERPEN

Estimating the Yield Curve Using the Nelson-Siegel Model

A Ridge Regression Approach

Jan Annaert

Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerp, Belgium

Anouk G.P. Claes

Louvain School of Management, Boulevard du Jardin Botanique 43, 1000 Brussels, Belgium

Marc J.K. De Ceuster°

Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerp, Belgium

Hairui Zhang

Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerp, Belgium

Abstract

The Nelson-Siegel model is widely used in practice for fitting the term structure of interest rates. Due to the ease in linearizing the model, a grid search or an OLS approach using a fixed shape parameter are popular estimation procedures. The estimated parameters, however, have been reported (1) to behave erratically over time, and (2) to have relatively large variances. We show that the Nelson-Siegel model can become heavily collinear depending on the estimated/fixed shape parameter. A simple procedure based on ridge regression can remedy the reported problems significantly.

Keywords: Smoothed Bootstrap, Ridge Regression, Nelson-Siegel, Spot Rates

° Corresponding Author – Email: marc.deceuster@ua.ac.be

1. Introduction

Good estimates of the term structure of interest rates (also known as the spot rate curve or the zero bond yield curve) are of the utmost importance to investors and policy makers. One of the term structure estimation methods, initiated by Bliss and Fama (1987), is the smoothed bootstrap. Bliss and Fama (1987) bootstrap discrete spot rates from market data and then fit a smooth and continuous curve to the data. Although various curve fitting spline methods have been introduced (quadratic and cubic splines (McCulloch (1971, 1975)), exponential splines (Vasicek and Fong (1982)), B-splines (Shea (1984) and Steeley (1991)), quartic maximum smoothness splines (Adams and Van Deventer (1994)) and penalty function based splines (Fisher, Nychka and Zervos (1994), Waggoner (1997)), these methods have been criticized on the one hand for having undesirable economic properties and on the other hand for being 'black box' models (Seber and Wild (2003)). Nelson and Siegel (1987) and Svensson (1994, 1996) therefore suggested parametric curves that are flexible enough to describe a whole family of observed term structure shapes. These models are parsimonious, they are consistent with a factor interpretation of the term structure (Litterman and Scheinkman (1991)) and they have both been widely used in academia and in practice. In addition to the level, slope and curvature factors present in the Nelson-Siegel model, the Svensson model contains a second hump/trough factor which allows for an even broader and more complicated range of term structure shapes. In this paper, we restrict ourselves to the Nelson-Siegel model. The Svensson model shares – by definition – all the reported problems of the Nelson-Siegel approach. Since the source of the problems, i.e. collinearity, is the same for both models, the reported estimation problems of the Svensson model may be reduced analogously.

The Nelson-Siegel model is extensively used by central banks and monetary policy makers (Bank of International Settlements (2005), European Central Bank (2008)). Fixed-income portfolio managers use the model to immunize their portfolios (Barrett, Gosnell and Heuson (1995) and Hodges and Parekh (2006)) and recently, the Nelson-Siegel model also regained popularity in academic research. Dullmann and Uhrig-Homburg (2000) use the Nelson-Siegel model to describe the yield curves of Deutsche Mark-denominated bonds to calculate the risk structure of interest rates. Fabozzi, Martellini and Priaulet (2005) and Diebold and Li (2006) benchmarked Nelson-Siegel forecasts against other models in term structure forecasts, and they found it performs well, especially for longer forecast horizons. Martellini and Meyfredi (2007) use the Nelson-Siegel approach to calibrate the yield curves and estimate the value-at-risk for fixed-income portfolios. Finally, the Nelson-Siegel model estimates are also used as an input for affine term structure models. Coroneo, Nyholm and Vidava-Koleva (2008) test to which degree the Nelson-Siegel model approximates an arbitrage-free model. They first estimate the Nelson-Siegel model and then use the estimates to construct interest rate term structures as an input for arbitrage-free affine term structure models. They find that the parameters obtained from the Nelson-Siegel model are not statistically different from those obtained from the 'pure' no-arbitrage affine-term structure models.

Notwithstanding its economic appeal, the Nelson-Siegel model is highly nonlinear which causes many users to report estimation problems. Nelson and Siegel (1987) transformed the nonlinear estimation problem into a simple linear problem, by fixing the shape parameter that causes the nonlinearity. In order to obtain parameter estimates, they computed the OLS estimates of a series of models conditional upon a grid of the fixed shape parameter. The estimates that, conditional upon a fixed shape parameter, maximized the R^2 were chosen. We refer to their procedure as a *grid search*. Others have suggested to estimate the Nelson-Siegel parameters simultaneously using *nonlinear optimization techniques*. Cairns and Pritchard (2001), however, show that the estimates of the Nelson-Siegel model are very sensitive to the starting values used in the optimization. Moreover, time series of the estimated coefficients have been documented to be very unstable (Barrett, Gosnell and Heuson (1995), Fabozzi, Martellini and Priaulet (2005), Diebold and Li (2006), Gurkaynak, Sack and Wright (2006), de Pooter (2007)) and even to generate negative long term rates, thereby clearly violating any economic intuition. Finally, the standard errors on the estimated coefficients, though seldom reported, are large.

Although these estimation problems have been recognized before, it has never lead towards satisfactory solutions. Instead, it became common practice to fix the shape parameter over the whole time series of term structures.[1] Hurn, Lindsay and Pavlov (2005), however, point out that the Nelson-Siegel model is very sensitive to the choice of this shape parameter. de Pooter (2007) confirms this finding and shows that with different fixed shape parameters, the remaining parameter estimates can take extreme values. Hence fixing the shape parameter is a non-trivial issue. In this paper we use ridge regression to alleviate the observed problems substantially and to estimate the shape parameter freely.

The remainder of this paper is organized as follows. In Section 2, we introduce the Nelson-Siegel model. Section 3 presents the estimation procedures used in the literature, illustrates the multicollinearity issue which is conditional on the estimated (or fixed) shape parameter and proposes an adjusted procedure based on the ridge regression. In the subsequent section (Section 4) we present our data and their

[1] Barrett et al. (1995) and Fabozzi et al. (2005) fix this shape parameter to 3 for annualized returns. Diebold and Li (2006) choose an annualized fixed shape parameter of 1.37 to ensure stability of parameter estimation.

descriptive statistics. Since the ridge regression introduces a bias in order to avoid multicollinearity, we will mainly evaluate the merits of the models based on their ability to forecast the short and long end of the term structure. The estimation results and the robustness of our ridge regression are discussed in Section 5. Finally, we conclude.

2. A first look at the Nelson-Siegel model

In their model Nelson and Siegel (1987) specify the forward rate curve $f(\tau)$ as follows:

$$f(\tau)=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}'\begin{bmatrix}1\\ e^{-\tau/\lambda}\\ (\tau/\lambda)e^{-\tau/\lambda}\end{bmatrix}=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}'\begin{bmatrix}f_0\\ f_1\\ f_2\end{bmatrix}, \tag{2.1}$$

where τ is time to maturity, β_0, β_1, β_2 and λ are coefficients, with $\lambda > 0$.

This model consists of three parts reflecting three factors: a constant (f_0), an exponential decay function (f_1) and a Laguerre function (f_2). The constant represents the (long-term) interest rate level. The exponential decay function reflects the second factor, a downward ($\beta_1 > 0$) or upward ($\beta_1 < 0$) slope. The Laguerre function in the form of xe^{-x}, is the product of an exponential with a polynomial. Nelson and Siegel (1987) chose a first degree polynomial which makes the Laguerre function in the Nelson-Siegel model generate a hump ($\beta_2 > 0$) or a trough ($\beta_2 < 0$). The higher the absolute value of β_2, the more pronounced the hump/trough is. The coefficient λ, referred to as the shape parameter, determines both the steepness of the slope factor and the location of the maximum (resp. minimum) of the Laguerre function.

The spot rate function, which is the average of the forward rate curve up to time to maturity τ, is defined as:

$$r(\tau)=\frac{1}{\tau}\int_0^\tau f(u)du, \tag{2.2}$$

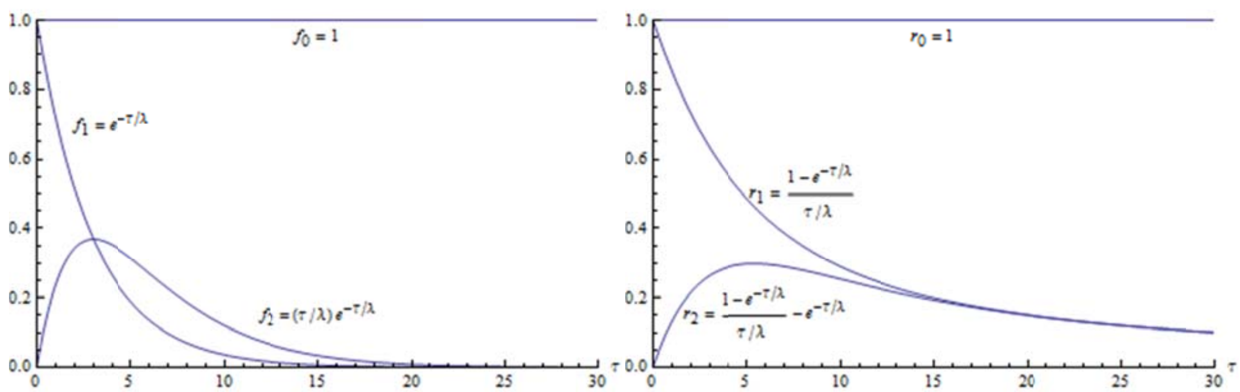
with continuous compounding. Hence, the corresponding spot rate function at time to maturity τ reads

$$r(\tau)=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}'\begin{bmatrix}1\\ \lambda\left(1-e^{-\tau/\lambda}\right)/\tau\\ \lambda\left(1-e^{-\tau/\lambda}\right)/\tau-e^{-\tau/\lambda}\end{bmatrix}=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}'\begin{bmatrix}r_0\\ r_1\\ r_2\end{bmatrix}. \tag{2.3}$$

Figure 1 depicts the three building blocks of the Nelson-Siegel model. The curves f_0, f_1 and f_2 in Panel A (respectively r_0, r_1 and r_2 in Panel B) represent the level, slope and curvature components of the forward rate (the spot rate) curve.

Figure 1: Decomposition of the Nelson-Siegel Model with the Shape Parameter Fixed at 3

Panel A: For the Forward Rate Curve **Panel B: for the Spot Rate Curve**



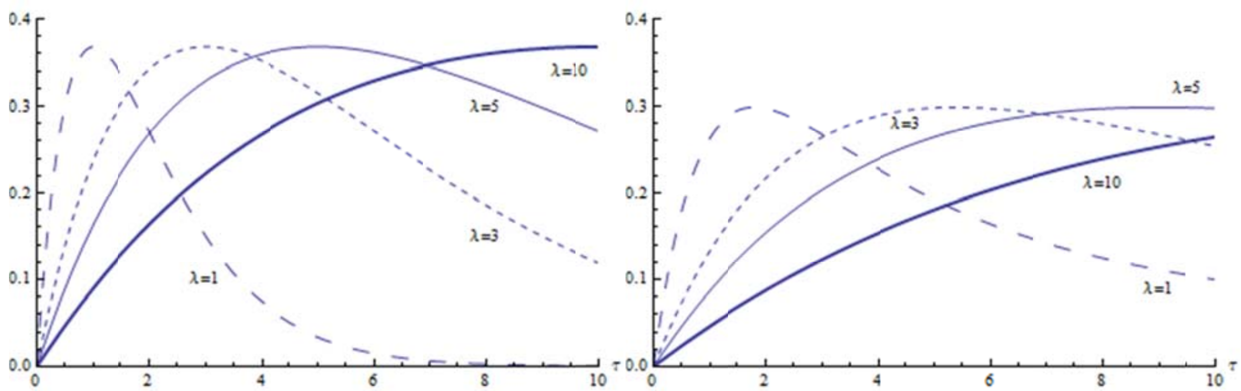
Note: This figure shows the decomposed components of the Nelson-Siegel model for the forward rate curve (Panel A) and the spot rate curve (Panel B) when the shape parameter is fixed at 3. The curves f_0, f_1 and f_2 in Panel A (respectively r_0, r_1 and r_2 in Panel B) represent the level, slope and curvature components of the forward rate (the spot rate) curve.

The role of the components becomes clear when we look at their limiting behaviour with respect to the time to maturity. When the time to maturity grows to infinity, the slope and curvature component vanish and the long-term forward and spot rate will converge to the same constant level of interest rate, β_0. Based on economic intuition, we assume β_0 to be close to the empirical long-term spot rate and not to be negative or unrealistically high. When the time to maturity approaches zero, only the curvature component vanishes and the forward and the spot rate converge to ($\beta_0 + \beta_1$). The spread, - β_1, measures the slope of the term structure, whereby a negative (positive) β_1 represents an upward (downward) slope. The degree of the curvature is controlled by β_2, the rate at which the slope and curvature component decay to zero. Finally, the location of the maximum/minimum value of curvature component is determined by λ. Note that λ determines both the shape of the curvature component and the hump/trough of the term structure. By maximizing the curvature component in the spot rate function with respect to λ, we are able to determine the location of the hump/trough of the term structure.

In Figure 2 the curvature component in the forward rate and the spot rate are depicted for shape parameters ranging from 1 to 10. The curvature component in the forward rate curve reaches its maximum when $\tau = \lambda$, whereas that in the spot rate curve reaches its maximum when $\tau > \lambda$, which is determined by simply maximizing r_2 in Equation 2.3 with λ fixed.

Figure 2: Shapes of the Curvature Component

Panel A : in the Forward Rate Curve **Panel B: in the Spot Rate Curve**



Note: This figure shows the shapes of the curvature component in the forward rate and the spot rate curve with respect to various shape parameters. The curvature component in the forward rate curve (Panel A) reaches its maximum when $\tau = \lambda$, whereas that in the spot rate curve (Panel B) reaches its maximum when $\tau > \lambda$, which is determined by simply maximizing r_2 in Equation 2.3 with λ fixed.

Alternatively, we can force the location of the hump/trough of the term structure to be at a given time to maturity, by fixing the shape parameter to a specific value. This also linearizes the model and hence facilitates estimation. Several authors use this idea. In their empirical work, Diebold and Li (2006) e.g. fix the maximum of the curvature component in the spot rate function at a maturity of 2.5 years. Thus by maximizing the curvature component in the spot rate function, the shape parameter of the hump/trough becomes approximately 1.37 with annualized data. In Fabozzi et al. (2005), where the shape parameter was set to be 3, the hump is located approximately at a maturity of 5.38 on the spot rate curve.

3. Estimation procedures

In order to obtain all the parameter estimates simultaneously, we could use nonlinear regression techniques. Ferguson and Raymar (1998) and Cairns and Pritchard (2001), however, show that the nonlinear estimators are extremely sensitive to the starting values used and that the probability of getting local optima is high. Taking these drawbacks into account, most researchers have fixed the shape parameter and have estimated a linearized version of the Nelson-Siegel model. The parameters of the Nelson-Siegel model have typically been estimated by minimizing the sum of squared errors (SSE) using (1) OLS over a grid of pre-specified λ's (Nelson and Siegel (1987)), and (2) a linear regression, conditional on a chosen fixed shape parameter λ (Diebold and Li (2006), de Pooter (2007), and Fabozzi et al. (2005)). We refer to these methods as the traditional measures. The estimated parameters using the traditional methods, however, are reported (1) to behave erratically in time, and (2) to have relatively large variances. We will first show that these problems result from multicollinearity problems. Next, we introduce a ridge regression approach to remedy the reported problems and how to judiciously fix the shape parameter in order to estimate the linearized Nelson-Siegel model.

3.1. The nature of the multicollinearity problem

Researchers have been aware of potential multicollinearity issues while estimating the Nelson-Siegel model. Diebold and Li (2006) e.g. indicate that the high correlation between the factors of the Nelson-Siegel (1987) model makes it difficult to estimate the parameters correctly. What seems to have gone unnoticed, however, is the fact that the correlation between the two regressors of the model depends on (the times to maturity of) the financial instruments chosen in the bootstrap. In order to illustrate this point, we consider four different sets of times to maturity:

1. 3 and 6 months, 1, 2, 3, 4, 5, 7, 10, 15, 20 and 30 years;
2. 3, 6, 9, 12, 15, 18, 21, 24, 30 months, 3-10 years;
3. 1 week, 1-12 month, and 2-10 years;
4. 1 week, 6 months, and 1-10 years.

The first set of maturities was used by Fabozzi et al. (2005). Diebold and Li (2006) opted for the second maturity vector. Since researchers are inclined to use all the data they can find, we study two extra vectors that also include additional shorter time to maturities.

Table 1 summarizes the correlations between the regressors for the λ values chosen by Diebold and Li (2006) and Fabozzi et al. (2005), using the four vectors of time to maturity that we consider. The table shows that the correlation between the slope and the curvature component of the Nelson-Siegel model heavily depends on the choice of the shape parameter. The second maturity vector e.g. implies correlations varying between -5% and -87% depending on the shape parameter chosen. The correlation also depends severely on the choice of the time to maturity vector. Using λ = 1.37, the correlation varies from -0.549 to 0.256, for the maturity vectors chosen. Setting λ = 3 produces correlations from -0.324 to -0.931. The vector containing the series of short maturities (the third vector) turns out to be the most sensitive to the collinearity issue.

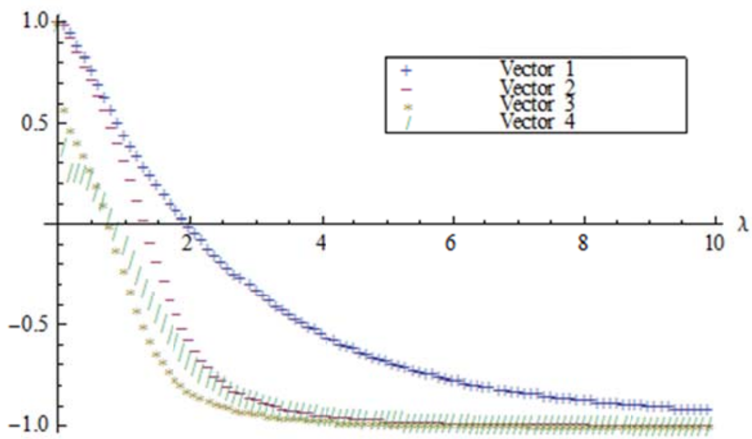
Table 1: Correlation between Regressors Using Alternative Time to Maturity Vectors

Maturity Vector	Diebold and Li *(λ = 1.37)*	Fabozzi et al. *(λ = 3)*
1	0.256	-0.324
2	-0.051	-0.871
3	-0.549	-0.931
4	-0.352	-0.872

Note: The correlations between the regressors in the Nelson-Siegel spot rate function conditioned on the shape parameter and the vector of times to maturity are presented. Vector 1 uses the following times to maturity: 3 and 6 months, 1, 2, 4, 5, 7, 10, 15, 20 and 30 years (as in Fabozzi et al. (2005)); vector 2 uses 3, 6, 9, 12, 15, 18, 21, 24, 30 months and 3-10 years (as in Diebold and Li (2006)); vector 3 is based on 1 week, 1-12 months and 2-10 years; and vector 4 on 1 week, 6 months and 1-10 years.

Figure 3 gives a more complete picture by plotting the correlation between the two regressors over a range of λ values using the four time to maturity vectors studied. We notice that the choice of the maturity vector influences the steepness of the correlation curve. It appears that Fabozzi et al. (2005) and Diebold and Li (2006) chose their λ values judiciously conditional on their maturity vector, although they both motivate their choice differently. It is clear that for empirical work it is of the utmost importance for all of the estimation methods to take this potential multicollinearity issue into account.

Figure 3: Correlations between the Slope and Curvature Component



Note: This figure plots the correlations between the slope and curvature component of the Nelson-Siegel model for the spot rate curve over the shape parameter λ using four time to maturity vectors. Vector 1 uses the following times to maturity: 3 and 6 months, 1, 2, 4, 5, 7, 10, 15, 20 and 30 years (as in Fabozzi et al. (2005)); vector 2 uses 3, 6, 9, 12, 15, 18, 21, 24, 30 months and 3-10 years (as in Diebold and Li (2006)); vector 3 is based on 1 week, 1-12 months and 2-10 years; and vector 4 on 1 week, 6 months and 1-10 years.

3.2 Traditional estimation methods

3.2.1 Grid search based OLS

To avoid nonlinear estimation procedures, Nelson and Siegel (1987) linearize their model by fixing λ and estimate Equation 2.3 with ordinary least squares. This procedure was repeated for a whole grid of λ values ranging from 0.027 to 1. The estimates with the highest R^2 were then chosen as the optimal parameter set. In practice, it is well known that grid search based OLS leads to parameter instability in the time series of estimates. This has been pointed out by many researchers including Barrett et al. (1995), Cairns et al. (2001), Fabozzi et al. (2005), Diebold et al. (2006 and 2008), Gurkaynak et al. (2006) and de Pooter (2007). Far less observed is that multicollinearity among the two regressors is the source of this instability. Moreover, high multicollinearity can also inflate the variance of the estimators.

3.2.2 OLS with fixed shape parameters

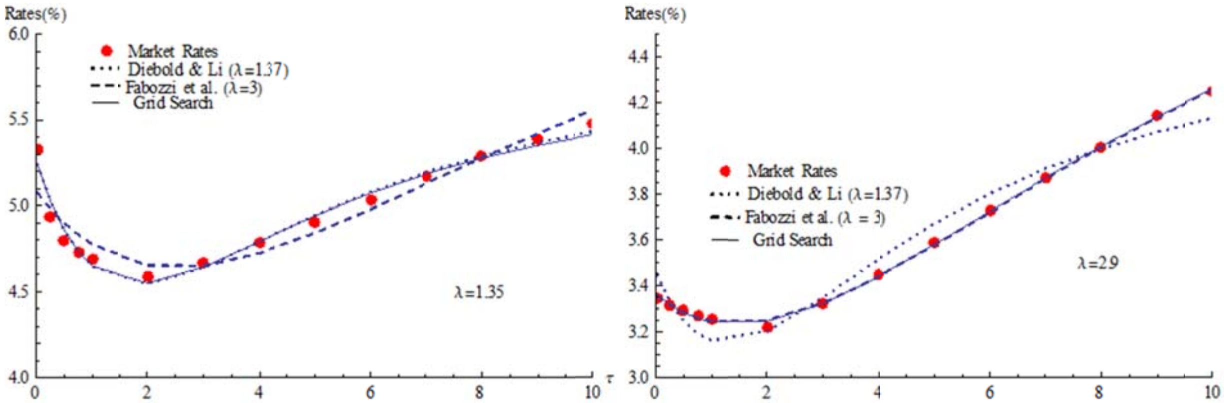
Some researchers fix the shape parameter which they typically motivate by prior knowledge about the curvature of the spot rates.

- Diebold and Li (2006) set λ to 16.4 with monthly compounded returns, or approximately 1.37 with annualized data. Their choice implies that the curvature component in the spot rate function will have its maximal value at the time to maturity of 2.5 years. They motivated their choice by stating that most of the humps/troughs are between the second and the third year. As we have shown, this choice, also turned out to avoid multicollinearity problems for their maturity vector.
- Fabozzi et al. (2005) fixed λ to 3 with annualized data, as argued by Barrett et al. (1995). Barrett et al. (1995) performed a grid search by fixing λ for the whole dataset and obtained a global optimal shape parameter in one go. In a footnote, Fabozzi et al. (2005) mention that when the shape parameter is fixed at 3, the correlation between the two regressors will not cause severe problems for their data. They, however, do not offer a universal procedure to tackle the multicollinearity issue.

From the previous section, we know that the degree of multicollinearity depends on the choice of the fixed shape parameter and the choice of the vector of times to maturity. The shape parameter that minimizes the squared errors may also vary over time. Figure 4 plots the spot rate curve on two days (i.e. April 18, 2001 and January 4, 1999) where fixing the optimal shape parameter to either 1.37 or 3 visually underperforms compared to the grid search. The λ values estimated using the grid search are reported in the figure. On the left hand side, the shape parameter is estimated to be 1.35. The use of a fixed λ of 1.37 will therefore result in a better fit than the model with a shape parameter fixed at 3. On the right hand side, however, the grid search estimate of λ is 2.9. The model using a fixed λ of 3 will thus fit the data better than that using a shape parameter of 1.37.

Figure 4: An Example of the Limitation of Fixed Shape Parameter

Panel A (Date: April 18, 2001) **Panel B (Date: January 4, 1999)**



Note: This figure gives an example of the limitation of fixed shape parameter. The grid search gives a shape parameter of 1.35 (2.9) which minimizes the fitting errors in Panel A (Panel B). As a consequence, the Nelson-Siegel model with a fixed shaped parameter of 1.37 (3) results in a better fit than the model with a fixed shaped parameter of 3 (1.37).

3.2.3 Grid search with conditional ridge regression

Whereas linear regressions do not require starting values for the estimators and always give globally optimal estimators, they do suffer from instability in parameter estimation. In this paper, we follow Nelson and Siegel's approach by combining the grid search with the OLS regression to 'free' the shape parameter. Conditional on the λ that results in the highest R^2, the parameters are re-estimated using ridge regression whenever the degree of multicollinearity among the regressors is too high. We therefore need to test the degree of multicollinearity of the two factors. The measure we use is discussed below. Subsequently, we discuss the nature of ridge regression and present the implementation of the ridge regression for the Nelson-Siegel term structure estimation.

3.2.3.1 Measuring the degree of multicollinearity

In order to address the multicollinearity issue, we need to verify the degree of collinearity. Popular collinearity measures include the variation inflation factor (VIF), the tolerance level and the condition number.

The VIF is defined as

$$\frac{1}{1-R_i^2},$$

where R_i^2 is the coefficient of multiple determination of the independent variable X_i on all other *X's* in the model. The intuition of VIF for a simple regression is straight forward: if the correlation between the regressors is high, then the VIF will be large. The tolerance level is the reciprocal of VIF.

Belsley (1991) points out the main drawbacks of these two measures: (1) high VIF's are sufficient but not necessary to collinearity problem, and (2) it is impossible to determine which regressors are nearly dependent on each other by using VIF's. The third measure, the condition number, is based on the eigenvalues of the regressors. Since the condition number avoids the shortcomings of the aforementioned methodologies (see Belsley (1991) and DeMaris (2004)), we use the condition number as the collinearity measure. Assume there is a standardized linear system $\boldsymbol{y} = \boldsymbol{\beta X} + \boldsymbol{\varepsilon}$. Denote κ (kappa) as the condition number and v the eigenvalues of $\boldsymbol{X'X}$. The condition number of $\boldsymbol{X}$ is defined as

$$\kappa(\mathbf{X}) = \frac{v_{\max}}{v_{\min}} \geq 1. \tag{3.1}$$

If $\boldsymbol{X}$ is well-conditioned (i.e. the regressor columns are uncorrelated), then the condition number is one, which implies that the variance is explained equally by all the regressors. If correlation exists, then the eigenvalues are no longer equal to 1. The difference between the maximum and minimum eigenvalues will grow as the collinearity effect increases. As suggested by Belsley (1991), we use a condition number of 10 as a measure of the degree of multicollinearity.[2]

3.2.3.2 Remedy of high collinearity

Once collinearity is detected, we need to remedy the problem. To overcome OLS parameter instability due to multicollinearity, we implement an alternative to the linear regression, i.e. the ridge regression technique. This estimation procedure can substantially reduce the sampling variance of the estimator, by adding a small bias to the estimator. Kutner, Nachtsheim, Neter and Li (2004) show that biased estimators with a small variance are preferable to the unbiased estimators with large variance, because

[2] As there are only two regressors in the Nelson-Siegel model, we can plot a one-to-one relationship between the condition number and the correlation between the two regressors. In our dataset, a condition number above 10 is equivalent to a correlation with an absolute value above 0.8.

the small variance estimators are less sensitive to measurement errors. We therefore use the ridge regression and compute our estimates as follows:

$$\widehat{\boldsymbol{\beta}}^{*} = \left[\mathbf{X}'\mathbf{X} + k\mathbf{I}\right]^{-1}\mathbf{X}'\mathbf{y}, \tag{3.2}$$

where k is called the ridge constant, which is a small positive constant. As the ridge constant increases, the bias grows and the estimator variance decreases, along with the condition number. Clearly, when k=0 the ridge regression is a simple OLS regression.

3.2.3.3 Implementation

As pointed out by Kutner et al. (2004), collinearity increases the variance of the estimators and makes the estimated parameters unstable. However, even under high collinearity, the OLS regression still generates the unbiased estimates. As a result, we implement a combination of the grid search and the ridge regression using the following steps:

1. Perform a grid search based on the OLS regression to obtain the estimate of λ which generates the lowest mean squared error.
2. Calculate the condition number for the 'optimal' λ.
3. Re-estimate the coefficients by using ridge regression only when the condition number is above a specific threshold (e.g. 10). The size of the ridge constant is chosen using an iterative searching procedure[3] that finds the lowest positive number, k, which makes the recomputed condition number fall below the threshold. By adding a small bias, the correlation between the regressors will decrease and so will the condition number.

4. Data and methodology

To illustrate our Nelson-Siegel term structure fitting procedure, we use Euribor rates maturing from 1 week up to 12 months and Euro swap rates with maturities between 2 years and 10 years. The Euro Overnight Index Average (EONIA), and the 20-, 25- and 30-year Euro swap rates were also collected to assess the out-of-sample prediction quality of the selected estimation procedures. The Euribor and

[3] We start with k = 0 and we iteratively re-compute the condition number after increasing the ridge constant with 0.001. We stop iterating when the recomputed condition number is lower than the pre-specified threshold (i.e. 10).

EONIA rates were obtained from their official website, and the Euro swap rates were gathered from Thompson DataStream®. Our dataset spans the period from January 4, 1999 to May 12, 2009 and includes 2644 days.

We use the smoothed bootstrap to construct the spot rate curves:

1. As the Euribor rates, $R'(\tau)$, with maturities less than one year use simple interest rates with the actual/360 day count convention, we convert them to

$$R(\tau) = \left[1 + R'(\tau) \times Days_{actual} / 360\right]^{365/Days_{Actual}}, \tag{4.1}$$

   where $R(\tau)$ is the annually compounded Euribor rate using actual/365 date convention.

2. Swap rates are par yields, so we bootstrap the zeros rates. Denote $S(\tau)$ as the swap rate, time to maturity being τ. The following equation helps us to extract the spot rates from the swap rates:

$$R(\tau) = \left[1 - S(\tau) \sum_{j=1}^{\tau} \frac{1}{\left[1 + R(j)\right]^{j}}\right]^{-1/\tau} - 1. \tag{4.2}$$

   Here τ = 2, ..., 10 and the Euro swap rates use the actual/365 day count convention.

3. As the relationship between Equation 2.1 and 2.3 only holds for continuously compounded rates, we need to convert the annualized spot rates to continuously compounded rates:

$$r(\tau) = \log\left[1 + R(\tau)\right]. \tag{4.3}$$

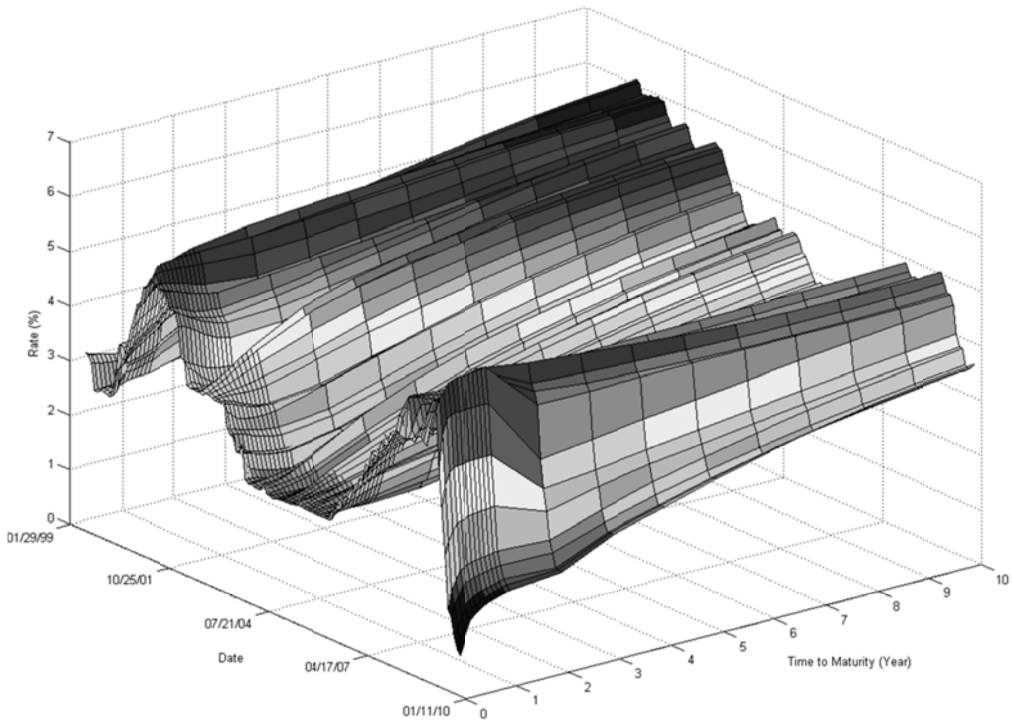
Table 2 summarizes the descriptive statistics for the time series of continuously compounded spot rates we use to fit the spot rate curve. The table shows that the volatility of the time series increases from 0.98% for weekly rates to 1.03% for 3-month rates, and then goes down to 0.69% for 10-year spot rate. The average spot rate increases as time to maturity grows, from 3.16% for the one week rate, to 4.53% for a 10-year maturity. Autocorrelation is high for rates of all maturities, from above 0.998 with a 5-day lag to above 0.915 with a 255-day lag.

Table 2: Descriptive Statistics of Spot Rates (Rates Are in Percentage)

Maturity	Mean	Std. Dev.	Min.	Max.	$\hat{\rho}(5)$	$\hat{\rho}(25)$	$\hat{\rho}(255)$
1 Week	3.186	0.982	0.710	5.240	0.998	0.972	0.926
1 Month	3.238	0.995	0.866	5.257	0.999	0.975	0.926
2 Months	3.292	1.013	1.111	5.299	0.999	0.980	0.935
3 Months	3.334	1.030	1.307	5.431	0.999	0.983	0.941
4 Months	3.351	1.027	1.379	5.441	0.999	0.984	0.944
5 Months	3.364	1.026	1.435	5.442	0.999	0.984	0.944
6 Months	3.377	1.025	1.501	5.449	0.999	0.984	0.945
7 Months	3.388	1.022	1.535	5.448	0.999	0.984	0.945
8 Months	3.400	1.020	1.566	5.445	0.999	0.984	0.945
9 Months	3.413	1.019	1.590	5.448	0.999	0.984	0.944
10 Months	3.426	1.017	1.613	5.444	0.999	0.984	0.943
11 Months	3.439	1.015	1.636	5.440	0.999	0.983	0.943
12 Months	3.453	1.014	1.656	5.451	0.999	0.983	0.942
2 Years	3.568	0.911	1.724	5.435	0.998	0.974	0.923
3 Years	3.742	0.845	2.160	5.513	0.998	0.972	0.918
4 Years	3.895	0.796	2.383	5.563	0.998	0.970	0.915
5 Years	4.029	0.762	2.599	5.614	0.998	0.971	0.916
6 Years	4.153	0.740	2.729	5.692	0.998	0.971	0.920
7 Years	4.267	0.726	2.845	5.755	0.998	0.973	0.925
8 Years	4.369	0.716	2.951	5.820	0.998	0.974	0.928
9 Years	4.456	0.705	3.047	5.891	0.998	0.975	0.932
10 Years	4.530	0.696	3.134	5.957	0.998	0.975	0.934

Note: Spot rates are expressed in percentage with continuous compounding. The sample period runs from January 4, 1999 to May 12, 2009, totaling to 2644 days. The spot rates with maturities less than one year are retrieved from the Euribor rates, whereas those with a maturity of more than one year are bootstrapped from Euro swap rates. Both are converted to obtain rates with according to the actual/365 (ISDA) date convention. $\hat{\rho}(n)$ is the n-day lag autocorrelation.

Figure 5: Time Series of the Euro Spot Rate Curves (1999 - 2009)



Note: This figure plots the time series of the Euro spot rate curve. The sample period runs from January 4, 1999 to May 12, 2009, totaling to 2644 days with the following times to maturity: 1 week, 1-12 months and 2-10 years. The spot rates with maturities less than one year are retrieved from the Euribor rates, whereas those with a maturity of more than one year are bootstrapped from Euro swap rates. Both are converted to obtain rates with according to the actual/365 (ISDA) date convention.

Figure 5 plots the time series of monthly spot rates in 3 dimensions. We can clearly observe that our data set contains a lot of variation in the level, the slope and the curvature of the term structure. The short-term spot rates (1 week) vary from approximately 3.75% to 5% in 2008, and decreases to almost 0.7% in 2009, due to the financial crisis. The long-term spot rate is relatively stable, varying between 3.134% and almost 6%. The yield curve is most often upward sloping. Around 2006 there are humps in the yield curves, while at other times there are troughs. Between 2003 and 2005 the yield curve is flatter compared to the S-shaped curves in other periods.

5. Empirical Comparison of the Estimation Methods

In order to compare the different estimation methods, we evaluate the estimation procedures based on the mean absolute error of their forecasting performance (the Mean Absolute Prediction Error or in short the MAPE). The absolute error is measured in basis points which gives us a good indication of the economic importance of the results. For every day in our time series we estimate the Nelson-Siegel model based on the proposed estimation methods, i.e. for the grid search, the OLS with the fixed shape parameter and the grid search using the conditional ridge regression. We then use the estimated term structures to forecast the spot rates used in the estimation (in-sample forecasting) and the contemporaneous EONIA, 20-, 25- and 30-year Euro swap rates (out-of-sample forecasting). The estimation procedure which produces (over the available time series) the lowest MAPE's 'wins' the rat race.

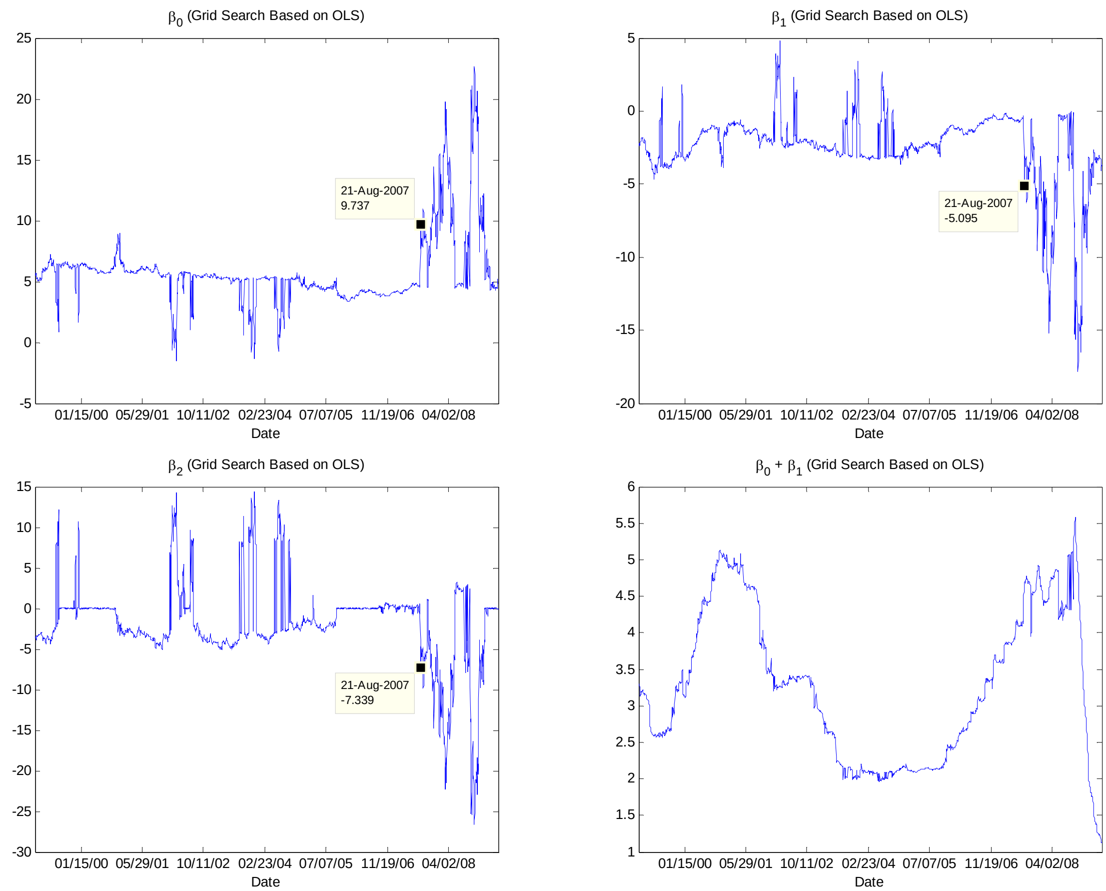
5.1 The Time Series of the Estimated Parameters

First, we discuss the results for the grid search, and subsequently we comment on the parameter estimates for the OLS approach where the shape parameter is fixed. Finally, we present the parameters for the grid search using the conditional ridge regression.

5.1.1 The grid search

Figure 6 graphically represents the time series of β_0, β_1, β_2 and ($\beta_0 + \beta_1$) estimates, based on a grid search using OLS. At some points in time, the three coefficients are clearly quite erratic. Moreover, for the long term interest rate level β_0, some negative values are obtained, thereby clearly violating any economic intuition. The short end of the term structure, denoted by ($\beta_0 + \beta_1$), is always positive. This is explained by highly negative correlation between the time series of β_0 and β_1 estimates. For August 21, 2007 e.g. the high β_0 coefficient (9.737) is accompanied with a low β_0 coefficient (-5.096) which leads to an estimate of the short rate of 5.477%.

Figure 6: Time Series of Estimated Parameters with the Grid Search Based on OLS

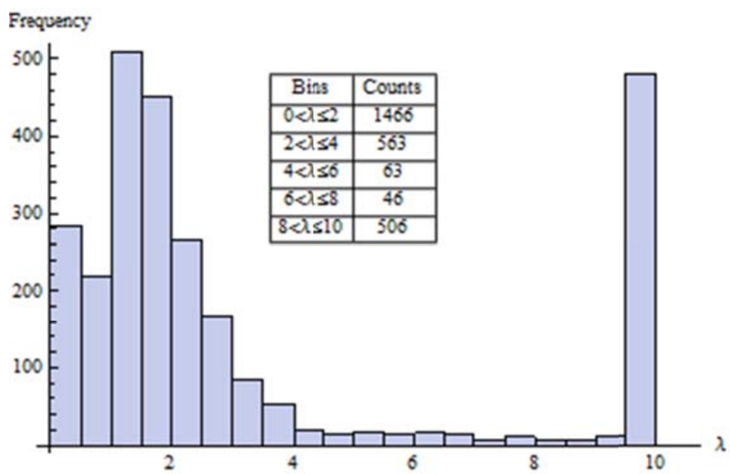


Note: This figure plots the time series of β_0, β_1, β_2 and ($\beta_0 + \beta_1$) estimates of the Nelson-Siegel model on Euro spot rates based on a grid search using OLS over the period from January 4, 1999 to May 12, 2009.

In an attempt to understand the source of the erratic time series behaviour, we report the histogram of the estimated shape parameters (Figure 7). The variation of the shape parameter estimates indicates that this parameter cannot be assumed constant over time. This confirms the visual inspection of the time series plot of our dataset (Figure 5) which revealed that the shape and the position of the humps changed over time. More than 55% of the estimated shape parameters are located within the range of 0 to 2, approximately 20% within the range of 2 to 4, and a little more than 19% within the range of 8 to

10. For the shape parameters within the range of 8 to 10, 472 out of 506 are estimated at the upper bound of the search interval i.e. at 10.[4]

Figure 7: Histogram of Optimal Shape Parameters Using Grid Search Based on OLS



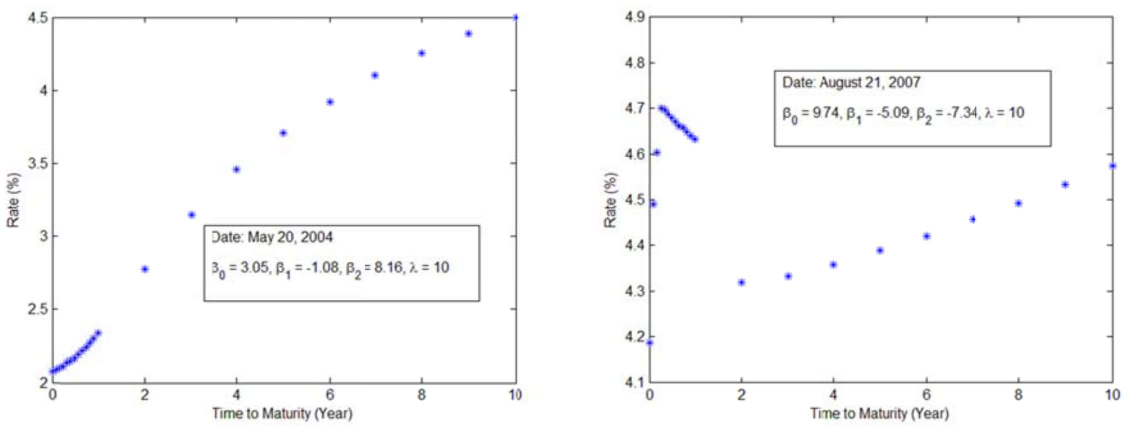
Note: This figure shows the histogram the estimated shape parameters λ of the Nelson-Siegel model on Euro spot rates using grid search based on OLS over the sample period from January 4, 1999 to May 12, 2009, totaling to 2644 days. Grid search is performed with the shape parameter ranging from 0 to 10.

There are two explanations for the relatively large amount of shape parameter estimates at the upper bound of the search interval: the absence of a hump/trough or the presence of more than one hump/trough. Both situations occur in our dataset as in shown in Figure 8.

[4] We performed grid search with the shape parameter ranging from 0 to 10, 0 to 20 (not reported) or 0 to 30 (not reported). The shape parameters that are estimated to be at the upper bound (e.g. 10) are also estimated at the upper bound of the search interval when this upper bound is 20 or 30.

Figure 8: The Term Structure of Spot Rates on May 20, 2004 (Panel A) and August 21, 2007 (Panel B)

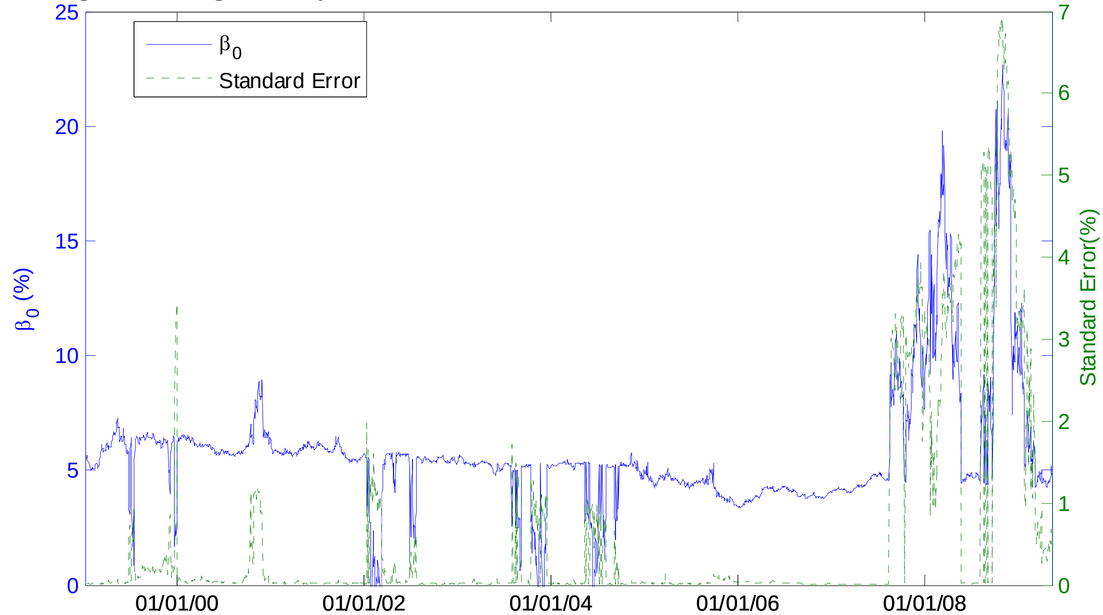
Panel A (Date: May 20, 2004) **Panel B (Date: August 21, 2007)**



Note: This figure depicts the term structure of spot rates on May 20, 2004 (Panel A) and August 21, 2007 (Panel B). The shape parameters based on grid search in both Panel A and Panel B are binding at 10. In Panel A the term structure does not show clear evidence of hump/trough. In Panel B the term structure has one hump and one trough, but the Nelson-Siegel model is not flexible enough to capture the shape of this term structure.

Figure 8 plots the term structure on May 20, 2004 and on August 21, 2007 with a shape parameter estimate at the upper bound of the search interval. In panel A of Figure 8, the situation of May 20, 2004 is presented. Since the term structure does not show a hump/trough, the shape parameter is estimated to be at the upper bound of the search interval. The optimization procedure therefore estimates the hump at the very end of the term structure. Figure 8, Panel B depicts the term structure of August 21, 2007. Here the term structure is too complicated to be described by the Nelson-Siegel model. Graphical inspection shows that one hump is not sufficient to closely fit the term structure on this day. Again, the optimal shape parameter is computed to be 10. Both examples illustrate problems resulting in upper bound shape parameter estimates. In the former case, the curvature component can simply be dropped from the Nelson-Siegel specification to fit the curve, with the additional benefit of eliminating of potential multicollinearity. In the latter case, a more flexible model such as Svensson (1994) will be a more appropriate candidate to describe the term structure.

Figure 9: Long Term Spot Rate Estimates and their Standard Errors – Grid Search



Note: This figure plots the long term spot rate estimates (solid line) and their standard errors (dashed line) based on grid search over the sample period from January 4, 1999 to May 12, 2009, totaling to 2644 days. The erratic behaviour of the Nelson-Siegel model based on grid search is illustrated.

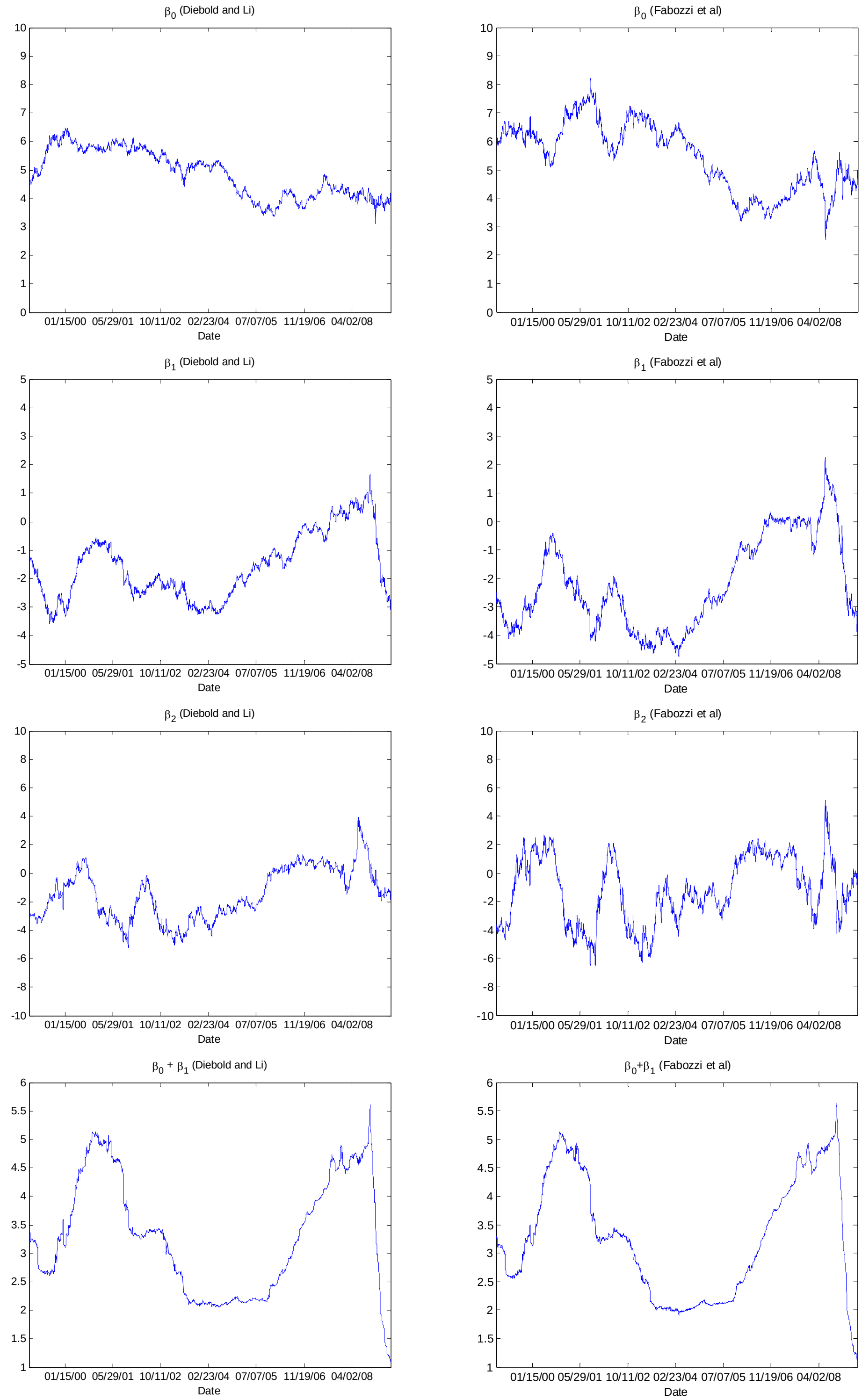
Grid search not only results in erratic time series of factor estimates, the precision of the estimates is also very time varying. Figure 9 redraws (as an example) the estimates of the long term spot rate (solid line) and its standard errors (dashed line). Whereas the standard errors are small at times, many periods of turbulence are shown in which the standard errors become 1% and more!

5.1.2 The OLS with fixed shape parameter

Figure 10 plots the time series of the estimated parameters conditional on a fixed shape parameter. In the left column we present the estimates using a fixed shape parameter of 1.37 (as in Diebold and Li (2006)), whereas on the right side those with a fixed shape parameter of 3 (as in Fabozzi et al. (2005)) are depicted. Generally, the time series of the estimates are not as volatile as those based on the grid search and the time series look smoothed. The higher shape parameter, however, results in a less smoothed time series of the estimated β_0, β_1 and β_2. The long-term interest rate level implied by the Nelson-Siegel model is always positive, and the short end does not show negative interest rate estimates either. However, the economic interpretation of the coefficients as factors, remains problematic. At the start of our series e.g., the long term rate is estimated as being 4.53% (λ = 1.37) and 5.74% (λ = 3) whereas the 30 year swap rate was 4.99%. So, which shape parameter should we consider to be the most appropriate?

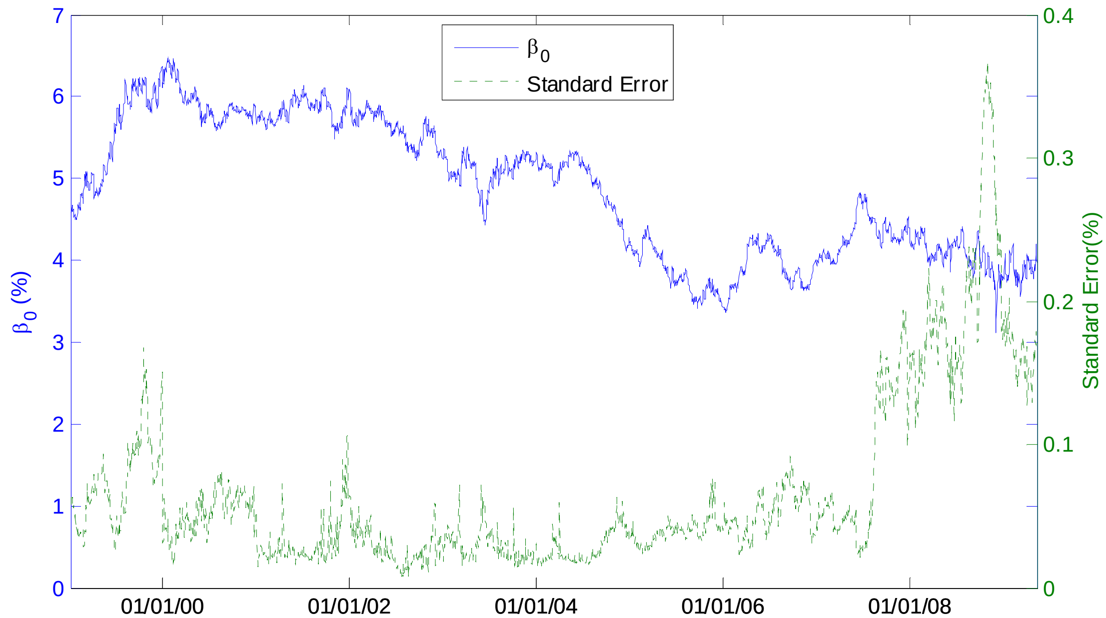
The precision of the estimates improves dramatically compared to the grid search. Figure 11 shows the standard errors for the long term rate using λ = 1.37. During the financial crisis, the standard error went up to more than 35 basis points. This is a serious improvement, compared to the standard errors of the grid search, where the standard errors went up to 500 basis points and more. Whether the shape parameter can best be fixed to 1.37 or to 3, however, remains an open question. Especially the results for β_2 are quite different for both choices in shape parameter, as can be seen in Figure 10. Our estimates also suggest that the economic characteristics of the time series of the estimated coefficients may be quite different. Recall that the Nelson-Siegel model can be interpreted in terms of a three factor model, the estimated coefficients being weights. Taking the variability of the shape parameter estimates we obtained in our grid search into account, it can be questioned whether the time variation in λ can be ignored at all!

Figure 10: Time Series of Estimated Parameters with Fixed Shape Parameters



Note: This figure plots the time series of β_0, β_1, β_2 and $(\beta_0 + \beta_1)$ estimates of the Nelson-Siegel model on Euro spot rates based on fixed shape parameters of 1.37 (Panel A) and 3 (Panel B). The sample period runs from January 4, 1999 to May 12, 2009.

Figure 11: Estimates of the Long Term Rate and their Standard Errors (*λ* = 1.37)

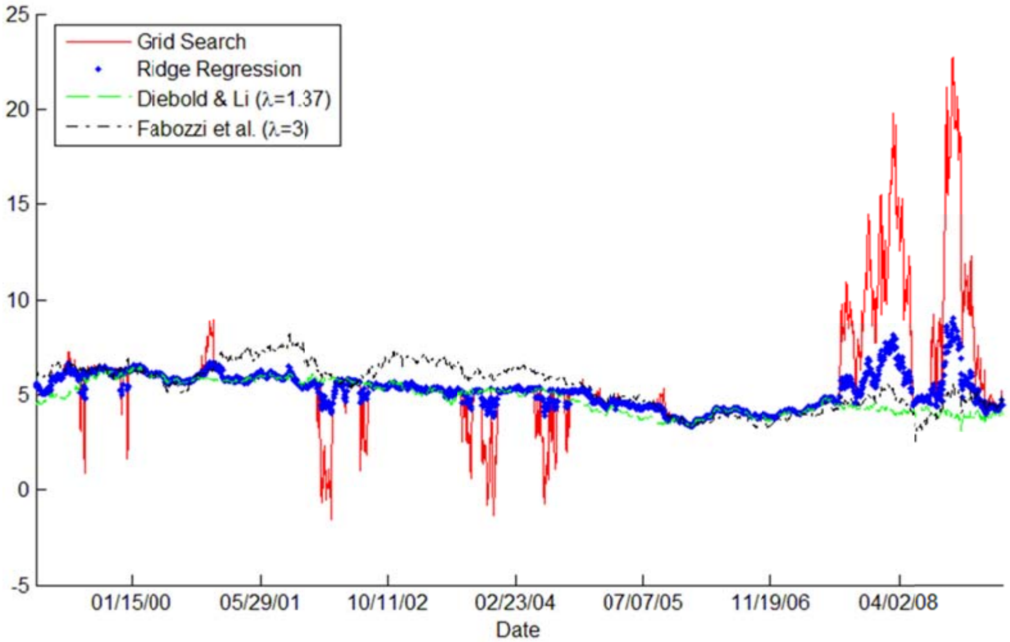


Note: This figure depicts the estimates of the long term rate (β_0) and their standard errors (Standard Error) based on a fixed shape parameter of 1.37 over the sample period from January 4, 1999 to May 12, 2009.

5.1.3 Grid search with conditional ridge regression

A drawback of the ridge regression technique is the lack of standard errors, which prohibits any kind of significance tests on the estimated coefficients (DeMaris (2004)). However, we can (visually) examine the stability of the time series estimates. Figure 12 again confirms that the time series of β_0, conditional on a fixed shape parameter of 3, almost perfectly coincides with the grid search β_0, time series. A low shape parameter smoothes the extreme jumps in the coefficients series almost completely, whereas the ridge regression takes a middle position. Whenever λ can be freely estimated (i.e. when λ is sufficiently low), no multicollinearity problems occur and the ridge regression is redundant. Whenever the correlation between the regressors exceeds our threshold, the ridge regression has a 'moderate' smoothing effect. However, the variation in the coefficients, e.g. β_0 in Figure 12 retains the strains put on the term structure.

Figure 12: The Estimated Long Term Rate Based on Alternative Estimation Methods

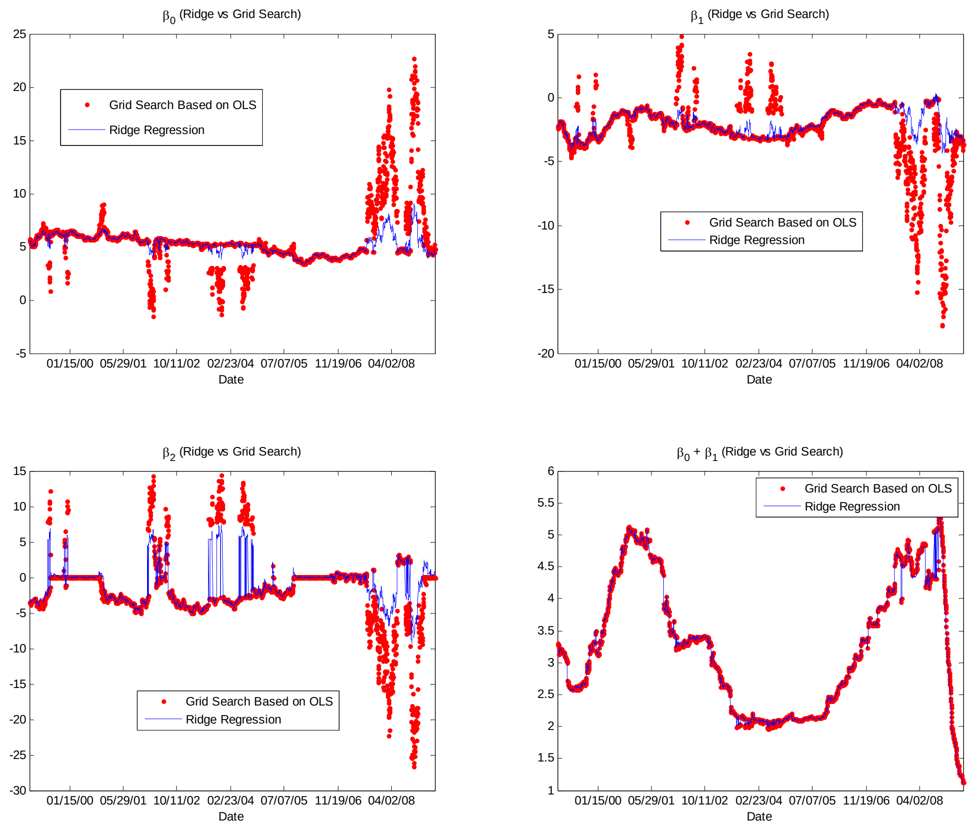


Note: This figure plots the estimated long term rates based on the following estimation methods: the grid search, the ridge regression, fixed shape parameter of 1.37 (Diebold and Li), and fixed shape parameter of 3 (Fabozzi et al.) over the sample period from January 4, 1999 to May 12, 2009.

Figure 13 shows the time series of all the estimated coefficients using the ridge regression, whereas on the right side the results from both the grid search based on OLS and the ridge regression are plotted.

The estimates from the ridge regression are more stable compared to the results from the grid search. There are no negative values in the long-term interest rate level anymore. The long-term interest rate level jumped up from 4.80% in June 2007 to 5.85% in August 2007, then 8.14% in March 2008, and then reached its peak of 8.99% in October 2008. Afterwards the long-term interest rate level went down. Ridge regression improves the stability of the estimates calculated by the grid search. And the positive long-term interest rate level complies with the economic intuition behind the Nelson-Siegel model. The short end of the term structure again is always positive, consistent with reality.

Figure 13: Estimated Parameters with the Ridge Regression and Grid Search



Note: This figure depicts the estimated parameters of the Nelson-Siegel model based on both grid search (line) and ridge regression (dot).β_0 represents the long term interest rate level implied by the Nelson-Siegel model. β_0+β_1 represents the short-term interest rate when time to maturity is zero.

In order to measure the stability of the time series of estimated coefficients more formally, we compute the standard deviation of their first differences (Table 3). We notice that ridge regressions have a lower volatility for all three parameters. Moreover, an *F*-test at the 95% confidence interval shows that the standard deviations of the ridge regression coefficients changes are significantly lower than those obtained through the grid search. The ridge regression, hence, can substantially reduce the instability of the estimates in the grid search. Compared with the fixed shape parameter procedures, the ridge regression allows the shape parameter to vary over time. Whether or not this is a desirable property from an economic point of view, still has to be seen. Therefore, we examine the in- and out-of-sample prediction performance of the various methods.

Table 3: The Standard Deviations of First Differences in the Estimates (in Percentage)

	Grid Search	Ridge Regression	λ =1.37	λ = 3
β_0	0.602	0.141	0.058	0.079
β_1	0.584	0.138	0.068	0.080
β_2	1.594	1.106	0.126	0.225

Note: This table shows the standard deviations of the first-order changes in the time series of the estimated parameters. The dataset used to estimate the parameters is composed of 1-week, 1- to 12-month, and 1- to 10-year spot rates. An *F*-test at a 95% confidence interval shows that all the standard deviations are significantly different from each other.

5.2 In-sample performance

In order to examine the in-sample performance, we compute the mean absolute errors between the predicted and the bootstrapped spot rate (Table 4). To test whether MAPE's are statistically different from each other, we compute:

$$\hat{\alpha}\bullet\mathbf{1} = \left|\mathbf{r} - \hat{\mathbf{r}}_{Method1}\right| - \left|\mathbf{r} - \hat{\mathbf{r}}_{Method2}\right|, \tag{5.1}$$

where ***r*** is the vector of empirical rates for a certain time to maturity, **1** is a vector of ones, $\hat{\mathbf{r}}_{Method1}$ and $\hat{\mathbf{r}}_{Method2}$ are the vectors of estimated spot rates using different methods (e.g. grid, ridge, Diebold and Li (DL) and Fabozzi et al. (FMP)). If the estimated coefficient $\hat{\alpha}$ is significantly negative, then we consider *Method 1* to be better than *Method 2*. The Newey-West correction is used to remove serial correlation from the residuals.

Although the overall MAPE is minimized by construction for the Grid Search, the in-sample MAPE's per maturity show how the fixed shape parameter methods perform compared to the ridge regression technique. The MAPE of FMP (DL) is for 13 (9) out of the 22 maturities the highest. For 8 out of the 22 maturities, the ridge regression generates the lowest MAPE, but it never produces the highest MAPE's. The grid search is for more than half of the maturities the MAPE minimizing method. The in-sample MAPE's reveal the limitations of reducing the model flexibility by fixing the shape parameters. Ridge regression outperforms these two methods convincingly.

Table 4: In-Sample Mean Absolute Prediction Errors

Maturity	Grid	Ridge	DL *(λ = 1.37)*	FMP *(λ = 3)*
1 Week	0.113	0.103*	0.130	0.170+
1 Month	0.067	0.059*	0.076	0.114+
2 Months	0.035	0.031*	0.033	0.062+
3 Months	0.031*	0.035	0.034	0.042+
4 Months	0.030*	0.034	0.036+	0.034
5 Months	0.031*	0.035	0.039+	0.036
6 Months	0.033*	0.035	0.043	0.047+
7 Months	0.033*	0.034	0.043	0.055+
8 Months	0.034*	0.034	0.044	0.063+
9 Months	0.035	0.035*	0.045	0.070+
10 Months	0.036	0.035*	0.045	0.075+
11 Months	0.038	0.037*	0.045	0.079+
12 Months	0.041	0.039*	0.046	0.083+
2 Years	0.055*	0.068	0.073+	0.057
3 Years	0.048*	0.065	0.077+	0.055
4 Years	0.033*	0.052	0.072+	0.065
5 Years	0.028*	0.044	0.065+	0.061
6 Years	0.024*	0.030	0.046	0.047+
7 Years	0.015	0.013*	0.017	0.026+
8 Years	0.013	0.018	0.022	0.008*
9 Years	0.020*	0.037	0.054+	0.034
10 Years	0.033*	0.055	0.083+	0.066

Note: In-sample mean absolute errors between produced data from the Nelson-Siegel and empirical data are presented. The dataset used to estimate the parameters is composed by 1-week, 1- to 12-month, and 1- to 10-year spot rates. * This approach yields the lowest MAPE for this time to maturity. + This approach yields the highest MAPE for this time to maturity. Except the following pairs, all the MAPE's are significantly different from each other (at 95% confidence interval) based on the *t*-test with Newey-West corrected standard errors: 2-month: grid and DL; 3-month: ridge and DL; 4-month: ridge and FMP; 5-month: ridge and FMP; 2-year: grid and FMP; 5-year: DL and FMP; 6-year: DL and FMP; 9-year: ridge and FMP. The mean absolute errors are tested using Equation 5.1.

5.3 Out-of-sample performance

Since the ridge regression adds a small bias to the OLS, we consider an out-of-sample performance test superior in order to judge the appropriateness of the various estimation methods. If we denote $\varepsilon_\tau = r_\tau - \widehat{r_\tau}$ as the residual between the spot rate r_τ observed from the market and the estimated spot rate by the Nelson-Siegel model $\widehat{r_\tau}$ for a certain time to maturity τ, the bias is estimated as the average of the ε_τ's. The out-of-sample bias on each day is measured by averaging the residuals after fitting the

EONIA, 20-, 25- and 30-year swap rates using four methods. Table 5 clearly shows that the bias introduced in the ridge regression does not affect the bias in the forecasted spot rates in a material way compared to the other estimation methods.

Table 5: Out-of-Sample Bias

	Grid	**Ridge**	**DL** ***(λ = 1.37)***	**FMP** ***(λ = 3)***
Mean	-0.070	4.00E-4	0.099	-0.089
Std. Dev.	0.384	0.152	0.102	0.214
Minimum	-2.216	-0.925	-0.356	-0.820
Maximum	0.703	0.369	0.393	0.558

Note: The bias is expressed in percentage. The sample period runs from January 4, 1999 to May 12, 2009, totaling to 2644 days for which the out-of-sample estimation bias is measured. The out-of-sample bias on each day is measured by averaging the residuals after fitting the EONIA, 20-, 25- and 30-year swap rates using four methods. All the biases are significantly different from each other based on a *t*-test with Newey-West corrected standard errors.

To investigate the ability of these four approaches to forecast the long and short end of the term structure, we will compare the estimated rates to the EONIA, the 20-, 25- and 30-year Euro swap rates. The MAPE will again be our criterion.

Theoretically speaking, the estimated $\beta_0 + \beta_1$ represents the short end of the term structure. However, the EONIA rates are the shortest-term spot rates that can be observed in the market. Consequently, if a model can predict the EONIA rates with the highest accuracy, it will be superior in predicting the short end of the yield curve.

The estimated EONIA rates are calculated by plugging time to maturity $\tau = 1/365$ into Equation 2.3 along with the estimated parameters for all four methodologies. Afterwards, the differences between the empirical and the estimated rates are examined following the same procedure as for the in-sample MAPE's tests.

In addition, we also check which model can best fit the long-term end of the term structure. However, long-term spot rates are not directly observable in the market either. Moreover, due to the lack of swap rates for certain times to maturity (e.g. 11- or 29-year swap rate) to bootstrap the long-term spot rate (e.g. 30-year spot rate), we will use the following procedure to check the forecasting ability:

1. Combining Equation 2.3 and 3.4, we calculate the estimated swap rates using

$$\hat{S}(\tau) = \frac{1 - \left[1 + \hat{R}(\tau)\right]^{-\tau}}{\sum_{j=1}^{\tau} \frac{1}{\left[1 + \hat{R}(j)\right]^{j}}}, \tag{5.2}$$

where τ = 2, ..., 30 are the times to maturity,

$$\hat{R}(\tau) = e^{\hat{r}(\tau)} - 1$$

is the estimated τ-year annually compounded spot rate, $\hat{r}(\tau)$ is the estimated continuously compounded spot rate, and $\hat{S}(\tau)$ is the estimated swap rate.

2. We calculate and test the MAPE's between the empirical and the estimated swap rates using the same procedure as previously.

The results are summarized in Table 6. The ridge regression always yields the statistically lowest MAPE's. The restriction DL and FMP put on the shape parameter makes them underperform compared to both the ridge and the OLS regression. Moreover, the ridge regression has lower prediction errors when predicting the long end, compared to the short end of the term structure.

Table 6: Out-of-Sample Mean Absolute Prediction Errors

Maturity	Grid	Ridge	DL *(λ = 1.37)*	FMP *(λ = 3)*
Overnight EONIA	0.254	0.246*	0.287+	0.278
20-year swap rate	0.201	0.142*	0.234+	0.221
25-year swap rate	0.262	0.150*	0.236	0.268+
30-year swap rate	0.304	0.152*	0.227	0.307+

Note: Out-of-sample MAPE's between produced data from Nelson-Siegel and empirical data are presented. The dataset used to estimate the parameters is composed by 1-week, 1- to 12-month, and 1- to 10-year spot rates. * This approach yields the lowest MAPE for this time to maturity. + This approach yields the highest MAPE for this time to maturity. Except the following pairs, all the MAPE's are significantly different from each other (at 95% confidence interval) based on the *t*-test with Newey-West correction on standard errors: 20-year: grid and FMP, DL and FMP; 25-year: DL and FMP, grid and FMP; 30-year: grid and FMP. The mean absolute errors are tested using Equation 5.1.

The superiority of the ridge regression procedure is not only statistically significant. Looking at the MAPE's for the 30 year swap rate, the MAPE's are lowered with 7 to 15 basis points. On the short end, the gain is more moderate, up to 4 basis points.

5.4 Robustness check on the forecasting ability

We consider two robustness checks to confirm the outperformance of the conditional ridge regression procedure we propose. First, we want to verify whether or not our results are mainly driven by the 2008-2009 financial crisis which is part of our dataset. Second, we have shown that the multicollinearity problem is severely affected by the choice of the maturity vector. We will examine whether our results are robust to the choice of a different maturity vector.

5.4.1 The impact of the financial crisis

As seen in Figure 13, the financial crisis has had a substantial impact on parameter estimation. We thus divide our dataset into two subsets, the pre-crisis period from January 4, 1999 to July 2, 2007 (2169 days), and the crisis period from July 3, 2007 to May 12, 2009 (475 days). The out-of-sample predicting power of all four approaches (Grid, Ridge, DL and FMP) is presented in Table 7 and Table 8.

A *t*-test between the pre-crisis and crisis period MAPE's at 95% confidence level shows that during the financial crisis the predicting ability of all four methods is significantly lowered. The predicting ability of the grid search drops dramatically for both the short and the long end of the term structure, while for the other methodologies, the financial crisis has more impact on the short end than on the long end.

Nevertheless, the ridge regression performs consistently in both periods, while the grid search clearly does not. During the financial crisis the grid search predictions are worst for the long-end of the yield curve whereas in the pre-crisis period, the FMP model is worst for these maturities.

The results shown in these two tables again confirm our previous findings that the ridge regression has superior predictability in forecasting both ends of the yield curve.

Table 7: Out-of-Sample Mean Absolute Prediction Errors (Pre-crisis Period)

Maturity	Grid	Ridge	DL *(λ = 1.37)*	FMP *(λ = 3)*
Overnight EONIA	0.161	0.161*	0.195+	0.169
20-year swap rate	0.142	0.137*	0.206	0.217+
25-year swap rate	0.161	0.140*	0.218	0.260+
30-year swap rate	0.169	0.131*	0.211	0.296+

Note: Out-of-sample MAPE's between produced data from Nelson-Siegel and empirical data are presented for the pre-crisis period from January 4, 1999 to July 2, 2007 (2169 days). The dataset used to estimate the parameters is composed by 1-week, 1- 12 month, and 1- to 10-year spot rates. * This approach yields the lowest MAPE for this time to maturity. + This approach yields the highest MAPE for this time to maturity. Except the following pairs, all the MAPE's are significantly different from each other based on the *t*-test with Newey-West correction on standard errors: EONIA: grid and ridge, grid and FMP, ridge and FMP; 20-year: grid and ridge, DL and FMP.

Table 8: Out-of-Sample Mean Absolute Prediction Errors (Crisis Period)

Maturity	Grid	Ridge	DL *(λ = 1.37)*	FMP *(λ = 3)*
Overnight EONIA	0.676	0.635*	0.704	0.775+
20-year swap rate	0.470+	0.165*	0.363	0.240
25-year swap rate	0.721+	0.198*	0.317	0.306
30-year swap rate	0.919+	0.251*	0.296	0.361

Note: Out-of-sample MAPE's between produced data from Nelson-Siegel and empirical data are presented for the crisis period from July 3, 2007 to August 12, 2009 (475 days). The dataset used to estimate the parameters is composed by 1-week, 1- 12 month, and 1- to 10-year spot rates. * This approach yields the lowest MAPE for this time to maturity. + This approach yields the highest MAPE for this time to maturity. Except the following pairs, all the MAPE's are significantly different from each other based on the *t*-test with Newey-West correction on standard errors: EONIA: grid and DL; 20-year: grid and DL; 25-year: DL and FMP; 30-year: ridge and DL, DL and FMP.

5.4.2 A Different Maturity Vector

In order to test the robustness of our results, we estimate the yield curves again using only 1-week, 6-month, 1 to 10-year spot rates. The results, shown in Table 9, also confirm our previous findings that the ridge regression has superior predictability in forecasting both ends for the yield curve. However, unlike for the other dataset, now the predictability of FMP on 25- and 30-year swap rates is not the worst, but the grid search is. Here the correlation between the slope and hump factor using the DL-estimation recipe is -0.35 instead of -0.55. The highest MAPE's for the 25- and 30-year swap rates obtained using the grid search based on the OLS regression, can be blamed on the instable estimates.

Table 9: Out-of-Sample Mean Absolute Prediction Errors

Maturity	Grid	Ridge	DL *(λ = 1.37)*	FMP *(λ = 3)*
Overnight EONIA	0.293	0.217*	0.255+	0.243
20-year swap rate	0.171	0.128*	0.205+	0.179
25-year swap rate	0.226+	0.140*	0.207	0.222
30-year swap rate	0.266+	0.147*	0.200	0.259

Note: Out-of-sample MAPE's between produced data from Nelson-Siegel and empirical data are presented. The dataset used to estimate the parameters is composed by 1-week, 6-month, and 1- to 10-year spot rates. * This approach yields the lowest MAPE for this time to maturity. + This approach yields the highest MAPE for this time to maturity. Except the following pairs, all the MAPE's are significantly different from each other based on the *t*-test with Newey-West correction on standard errors: EONIA: grid and DL; 20-year: grid and FMP; 25-year: grid and DL, grid and FMP, DL and FMP; 30-year: grid and DL.

Again, we test the influence of the financial crisis on the performance of the proposed methods, this time on our limited sample. The results from this analysis comparing estimates of the pre-crisis and crisis period are summarized in Table 10 and Table 11.

At the 95% confidence level, a *t*-test between the pre-crisis and crisis period MAPE's shows that the performance of the four methods behaves similar to that of the other dataset. The high volatility of the estimates in the grid search makes its performance drop substantially during the financial crisis.

The ridge regression always has the highest predicting ability except for the 30-year swap rate during the crisis period, where the DL has the highest predicting ability. However, the difference between these two methods for the 30-year swap rate during the financial crisis is not statistically significant.

Table 10: Out-of-Sample Mean Absolute Prediction Errors (Pre-crisis Period)

Maturity	Grid	Ridge	DL *(λ = 1.37)*	FMP *(λ = 3)*
Overnight EONIA	0.216+	0.156*	0.186	0.160
20-year swap rate	0.128	0.120*	0.191+	0.173
25-year swap rate	0.150	0.124*	0.201	0.209+
30-year swap rate	0.161	0.121*	0.192	0.242+

Note: Out-of-sample MAPE's between produced data from Nelson-Siegel and empirical data are presented for the pre-crisis period from January 4, 1999 to July 2, 2007 (2169 days). The dataset used to estimate the parameters is composed by 1-week, 6-month, and 1- to 10-year spot rates. * This approach yields the lowest MAPE for this time to maturity. + This approach yields the highest MAPE for this time to maturity. Except the following pairs, all the MAPE's are significantly different from each other based on the *t*-test with Newey-West correction on standard errors: 25-year: DL and FMP.

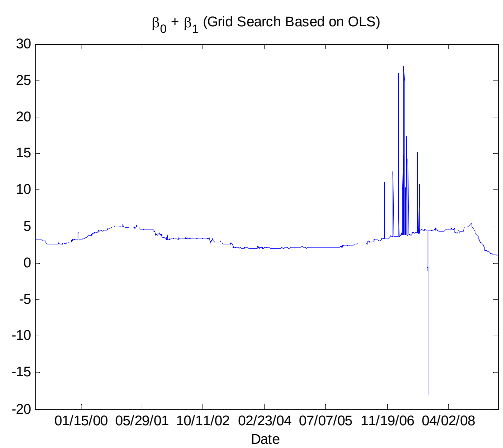
Table 11: Out-of-Sample Mean Absolute Prediction Errors (Crisis Period)

Maturity	Grid	Ridge	DL *(λ = 1.37)*	FMP *(λ = 3)*
Overnight EONIA	0.642+	0.499*	0.571	0.624
20-year swap rate	0.365+	0.164*	0.268	0.205
25-year swap rate	0.578+	0.213*	0.236	0.280
30-year swap rate	0.749+	0.264	0.236*	0.337

Note: Out-of-sample MAPE's between produced data from Nelson-Siegel and empirical data are presented for the crisis period from July 3, 2007 to August 12, 2009 (475 days). The dataset used to estimate the parameters is composed by 1-week, 6-month, and 1- to 10-year spot rates. * This approach yields the lowest MAPE for this time to maturity. + This approach yields the highest MAPE for this time to maturity. Except the following pairs, all the MAPE's are significantly different from each other based on the *t*-test with Newey-West correction on standard errors: EONIA: grid and DL, grid and FMP; 25-year: ridge and DL, DL and FMP; 30-year: ridge and FMP.

One thing worth mentioning is that for the alternative dataset, not only the long end of the term structure based on the grid search is extremely volatile and sometimes negative, so is the short end of the term structure, as shown in Figure 14. This again shows that the grid search is very sensitive to the underlying dataset.

Figure 14: The Short End of the Term Structure by the Grid Search Based on OLS



Note: This figure plots the short end of the term structure implied by the grid searched based on OLS over the sample period from January 4, 1999 to May 12, 2009, totaling to 2644 days.

6. Conclusion

Many researchers have reported problems in estimating the Nelson-Siegel (1987) model. We have shown that multicollinearity between the slope and hump factor are causing both instability of the regression coefficients over time and large standard errors on the coefficients. To temper the estimation problems, we apply ridge regression technique, whenever the grid search based estimate of the shape parameter results in highly correlated slope and hump factors. For euro spot rate curves, over the period 199-2009, we compare the grid search estimates - originally proposed by Nelson and Siegel - to our ridge regression estimates. The Diebold and Li (2006) and the Fabozzi et al. (2005) estimates were also calculated as a benchmark.

The in-sample comparison shows that the grid search produces erratic time series estimates which sometimes violate the economic intuition behind the Nelson-Siegel model. The distribution of the freely estimated shape parameters and the in-sample fitting errors reveal the limitation of the use of a fixed shape parameter. The out-of-sample predictability at the two ends of the term structure shows that the ridge regression always produces the lowest mean absolute prediction errors. For the long end of the term structure, the economic gain in the MAPE mounted up to 15 basis points. For the short end, the differences in MAPE's were statistically significant but economically smaller.

Robustness checks show that the ridge regression performs robustly and consistently better in different economic environments (pre-crisis and crisis period) and with different choices of the maturity vector (i.e. the set of financial instruments used to bootstrap the yield curve).

Based on our findings, fixing the shape parameter in order to avoid multicollinearity, is a statistical trick that does reduce the correlation between the regressors when the fixing is judicially chosen. It however ignores the bare fact that in practice, term structures do take all kind of (humpy) shapes and that the hump simply is not fixed over the time to maturity spectrum. The loss in flexibility comes at a severe price especially for the predictions of long term spot rates. This paper advances a procedure based on ridge regression to reduce the prediction errors significantly.

7. References

Adams, K. “Smooth Interpolation of Zero Rates”, Algo Research Quarterly, Vol. 4, 1-2, (2001) 11-22.

Adams, K. J. and Van Deventer D. R. “Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness”, Journal of Fixed Income, Vol. 4, No. 1 (June 1994), 52-62.

Bank of International Settlements. “Zero-Coupon Yield Curves – Technical Documentation”, BIS Paper No. 25, 2005.

Barrett, W. R., Gosnell, T. F. Jr. and Heuson, A. J., “Yield Curve Shifts and the Selection of Immunization Strategies”, Journal of Fixed Income, Vol. 5, No. 2 (September 1995), 53-64.

Belsley, D.A. “Conditioning Diagnostics – Collinearity and Weak Data in Regression”, Wiley Series in Probability and mathematical Statistics, 1991.

Bliss, R., and Fama, E., “The Information in Long-Maturity Forward Rates”, American Economic Review, Vol. 77 (1987), 680-692.

Cairns, A. J. G., and Pritchard, “Stability of Descriptive Models for the Term Structure of Interest Rates with Applications to German Market Data”, British Actuarial Journal Vol. 7 (2001), 467-507.

Causey, B. D., “A Frequentist analysis of a Class of Ridge Regression Estimators”, Journal of American Statistical Association, Vol. 75, No. 371 (September 1980), 736-738.

Clikff, A. D. and Ord, J. K., “Spatial Processes – Models & Applications”, Pion Limited, 1981.

Cohen, J., Cohen, P., West, S. G., and Aiken, L.S., “Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd Edition)”, Lawrence Erlbaum Associates, Publishers, 2003.

Coroneo, L., Nyholm, K., and Vivada-Koleva, R., “How Arbitrage-Free Is the Nelson-Siegel Model?”, European Central Bank Working Paper Series, No. 874 (February 2008).

de Pooter, M., “Examining the Nelson-Siegel Class of Term Structure Models”, Tinbergen Institute Discussion Paper, IT 2007-043/4 (2007).

DeMaris, A., “Regression with Social Data: Modeling Continuous and Limited Response Variables”, Wiley Series in Probability and Statistics 2004.

Diebold, F. X. and Li, C., "Forecasting the Term Structure of Government Bond Yields", Journal of Econometrics, Vol. 130 (2006), 337-364.

Diebold, F. X., Li, C., and Yue, V. Z., "Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach", Journal of Econometrics, Vol. 146 (2008), 351-363.

Dullmann, K. and Uhrig-Homburg, M. "Profiting Risk Structure of Interest Rates: An Empirical Analysis for Deutschemark-Denominated Bonds", European Financial Management, Vol. 6, No. 3 (2000), 367-388.

European Central Bank, The New Euro Area Yield Curves, Monthly Bulletin, (February 2008), 95-103.

Fabozzi, F. J., Martellini, L. And Priaulet, P., "Predictability in the Shape of the Term Structure of Interest Rates", Journal of Fixed Income, Vol. 15, No. 1 (June 2005) 40-53.

Fisher, M., Nychka, D., and Zervos, D. "Fitting the Term Structure of Interest Rates with Smoothing Splines", Finance and Economics Discussion Series, Federal Reserve Board, September, 1994.

Gurkaynak, R. S., Sack, B. and Wright, J. H., "The U.S. Treasury Yield Curve: 1961 to the Present", Working paper in the Finance and Economics Discussion Series (FEDS), Federal Reserve Board, Washington D.C. 2006-28, 2006.

Hurn, A.S., K.A. Lindsay, and V. Pavlov, "Smooth Estimation of Yield Curves by Laguerre Function", Queensland University of Technology Working Paper (2005), 1042-1048.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W., "Applied Linear Statistical Models (5th Edition)", McGraw-Hill, 2004.

Litterman, R. and J. Scheinkman, "Common Factors Affecting Bond Returns", Journal of Fixed Income, Vol. 1, No. 1 (June 1991), 54-61.

Martellini, L, and Meyfredi, J. C., "A Copula Approach to Value-at-Risk Estimation for Fixed-Income Portfolios", Journal of Fixed Income, Vol. 17, No. 1 (June 2007), 5-15.

Nelson, C., and Siegel, A. F., "Parsimonious Modeling of Yield Curves", Journal of Business, Vol. 60 (October 1987), 473-489.

Seber, G. A. F., and Wild, C. J., "Nonlinear Regression", Wiley Series in Probability and Statistics, 2003.

Steeley, J. M., “Estimating the Gilt-Edged Term Structure: Basis Splines and Confidence Intervals”, Journal of Business Finance & Accounting, Vol. 18, No. 4 (June 1991), 513-529.

Svensson, L. E. O., “Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994”, IMF Working Paper, WP/94/114 (September 1994), 1-49.

Svensson, L. E. O., “Estimating the Term Structure of Interest Rates for Monetary Policy Analysis”, Scandinavian Journal of Economics, Vol. 98 (1996), 163-183

Vasicek, O. A., and Fong, H. G., “Term Structure Modeling Using Exponential Splines”, Journal of Finance, Vol. 73 (May 1982), 339-348.