AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2018. **16**: 8-29 DOI:10.5605/IEB.16.1

On multicollinearity and the value of the shape parameter in the term structure Nelson-Siegel model

León, Angel
Rubia, Antonio
Sanchis-Marco, Lidia

▸ RECEIVED: 10 AUGUST 2016
▸ ACCEPTED: 1 DECEMBER 2016

Abstract

This paper investigates the sensitivity of the dynamic Nelson-Siegel factor loadings to the value of the shape parameter, λ. It also analyses the multicollinearity problem and addresses how to mitigate this issue in the estimation process. First, we find that the selection of a fixed λ is not optimal due to the collinearity problems. Second, we observe a substantial difference between the forecasting performance of the traditional estimation procedures and that of the ridge regression approach. Finally, we implement a Monte Carlo simulation exercise in order to study the statistical distribution of the estimates of the model parameters and thus determine the extent to which they differ from the real values. Furthermore, we find that multicollinearity between the factors of the NS model can, in the case of ordinary least squares estimation with a fixed parameter λ, result in greater differences between the estimates and the actual parameter values. Ridge regression corrects such differences and produces more stable estimates than the ordinary linear and nonlinear least squares methods.

Keywords:
Term structure of interest rates, Yield curve, Nelson-Siegel, Ridge regression, Forecasting.

JEL classification:
C5, C32, C53, E4.

Please cite this article as:
León, A., Rubia, A. and Sanchis-Marco, L. (2018). On multicollinearity and the value of the shape parameter in the term structure Nelson-Siegel model, *AESTIMATIO, The IEB International Journal of Finance*, 16, pp. 8-29. doi: 10.5605/IEB.16.1

León, A. Department of Quantitative Methods and Economic Theory. University of Alicante. E-mail: aleon@ua.es.
Rubia, A. Department of Financial Economics. University of Alicante. E-mail: antonio.rubia@ua.es.
Sanchis-Marco, L. ✉ Department of Economic Analysis and Finance. University of Castilla-La Mancha. CP 45071, Toledo, Spain. ✆ (34) 925 268 800. E-mail: lidia.sanchis@uclm.es.

AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2018. **16**: 8-29

Sobre multicolinealidad y el valor del **parámetro de forma en el modelo de estructura temporal de Nelson y Siegel**

León, Angel
Rubia, Antonio
Sanchis-Marco, Lidia

Resumen

Este artículo investiga la sensibilidad de las cargas factoriales del modelo dinámico de Nelson y Siegel al valor del parámetro de forma λ, y analiza el problema de la multicolinealidad y cómo mitigarlo en el proceso de estimación. En primer lugar, se obtiene que la selección de un λ fijo no conduce a la optimalidad debido a que pudiera dar lugar a problemas de multicolinealidad. En segundo lugar, se observa una diferencia sustancial en los resultados de predicción entre los procedimientos tradicionales de estimación y el método de regresión alomada (*ridge regression*). Finalmente, se implementa un ejercicio de simulación de Monte Carlo con el fin de estudiar la distribución estadística de de las estimaciones de los parámetros del modelo, para comprobar las diferencias respecto a los valores reales. Se observa que la multicolinealidad entre las cargas factoriales del modelo de NS puede dar lugar, en el caso de estimación mínimo cuadrática lineal con parámetro de forma fijo, a mayores diferencias entre las estimaciones y los valores reales de los parámetros del modelo. La regresión alomada corrige estas diferencias y da lugar a estimaciones más estables que los procedimientos de estimación, lineal o no lineal, mínimo cuadráticos ordinarios.

Palabras clave:

Estructura temporal de los tipos de interés, curva de tipos de interés, Nelson-Siegel, regresión alomada, predicción.

1. Introduction

The term structure of interest rates refers to the relationship between bonds of different maturities. The estimation and forecasting of the term structure are of great interest to both academics and practitioners, and gaining an understanding of what moves bond yield is important for a number of reasons. First, yield forecasts provide a basis for firms' investment decisions, consumers' savings decisions and also policy decisions. Second, a model of the yield curve provides an insight into how movements in the short term translate into longer-term yields, which is key when it comes to assessing the impact of monetary policy. Finally, the evolution of the yield curve is important for derivative pricing and hedging strategies. There are different methods in the literature to estimate the yield curve: the traditional no-arbitrage and affine equilibrium models (see, among others, Vasicek, 1977; Cox *et al*., 1985; Hull and White, 1990) and the factor models (Nelson and Siegel, 1987; Svensson, 1997; Jakas, 2011).

In this article, we focus on the factor models, and specifically on the Nelson-Siegel (NS hereafter) exponential components framework, which offers a number of advantages for term-structure forecasting. Public organizations, investment banks and central banks rely heavily on NS-type models to fit and forecast yield curves. In addition, according to the BIS (2005) and the European Central Bank (2008), a wide range of European central banks use this model for estimating zero-coupon yield curves. Furthermore, similar to BIS III requirements, the Solvency II Directive sets out new regulation that dictates a Solvency Capital Requirement (SCR) for the European insurance market. Solvency II also provides a standard framework for a wide variety of risk management purposes. Within the context of this new regulation, practitioners and academics identify the NS model as an optimal and appropriate specification to determine the capital requirement for interest rate risk (e.g., see Abeling, 2013), and several extensions of this model have been proposed in recent years. The NS model has grown more popular due to its parsimonious estimation using linear regression when the shape parameter has been fixed, its ability to estimate yields for all maturities, the intuitive interpretations of the factors obtained and, lastly, because of its good performance in out-of-sample forecasting, as shown by Diebold and Li (2006), De Pooter *et al*. (2007), Rezende and Ferreira, (2011) and Molenaars *et al*., (2013, 2015), among others.

Diebold and Li (2006) dinamized the yield curve model proposed by Nelson and Siegel, in terms of a state-space representation using a simple two-step approach and fixing the shape parameter, λ. More recent studies that implement this simple ordinary least squares (OLS) regression approach as an approximation to the initial nonlinear problem include Li and Yue (2008) and Annaert *et al*. (2010). Those studies acknowledge that the nonlinear estimation of the original NS model produces a number of problems, including the sensitivity of the estimates of the NS model parameters to the starting values used in the optimization process; the very unstable time series of the estimated coefficients, with large standard errors; negative long-term rates; and multicollinearity between the regressors (factor loadings) of the model (see Section 2). Although these problems had previously been reported, no satisfactory solutions can be found in the literature. The main limitation of most of the existent studies is that they largely overlook the impact of λ on the NS estimation resulting from its external determination. We try to fill this gap by analysing the effect of fixing this parameter on the results of different NS estimation procedures using both real and simulated yield data. First, we use real data to analyse the correlation between the factor loadings that depend on the λ-parameter, in order to establish how multicollinearity is conditional on this shape parameter. Next, we attempt to show that both the linear and nonlinear NS estimation procedures used in the literature are affected by the abovementioned problems (for specific values of the shape parameter), which can be avoided using ridge regression. Then, we study the out-of-sample forecasting performance of these three estimation procedures. Finally, using simulated data, we determine the main descriptives that characterize the statistical distribution of the parameters included in the NS model.

This analysis can be considered an extension of the very interesting recent work by Annaert *et al*. (2013). Their study concentrates only on the implementation of the ridge regression methodology using the Euro term structure database, without analysing the goodness of the estimators. We fill this gap by implementing an in-depth Monte Carlo simulation, something that has not been attempted in previous studies. Furthermore, alternative econometric techniques, such as nonlinear least squares, can be applied. In short, we can check how well the sample distribution of the estimated parameters agree with the real values, depending on the degree of multicollinearity in the regressors.

The main contribution of this paper is to demonstrate that the OLS estimation of the NS model is not suitable when the multicollinearity between the factor loadings is high. Ridge regression emerges as a possible alternative which produces both better estimates of the model parameters and better out-of-sample forecasting performance. A simulation exercise allows us to examine this and also to offer a more complete analysis of the distribution of the NS-estimated parameters for linear, nonlinear and ridge regression.

The remainder of this paper is organized as follows. In section 2, we present the NS Model. Section 3 sets out the NS estimation procedures we want to compare. Section 4 is devoted to the empirical analysis and includes the Diebold and Li (2006) database, the estimation results and the forecasting analysis. Section 5 addresses the

Monte Carlo simulation study used to analyse the behaviour of the NS-estimated parameters conditional on the λ value. Finally, we provide our conclusions in Section 6.

2. Modelling the term structure: Nelson and Siegel approach

The term structure of interest rates shows the relationship between the interest rates and maturities of zero-coupon bonds without risk of default. Let $P_t(\tau)$ denote the price of a τ-period discount bond, that is, the present value at time t of €1 receivable τ periods ahead

$$P_t(\tau) = e^{-\tau y_t(\tau)} \tag{1}$$

where $y_t(\tau)$ denotes the continuously compounded zero-coupon nominal yield to maturity τ. We focus on the Nelson and Siegel (1987) forward rate curve, which can be viewed as a constant plus a Laguerre function[1]. The yield curve is

$$y_t(\tau) = \beta_{1t} F_{1t} + \beta_{2t} F_{2t} + \beta_{3t} F_{3t} \tag{2}$$

where

$$F_{1t} = 1;\; F_{2t} = \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau};\; F_{3t} = \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \tag{3}$$

Therefore, Nelson and Siegel propose fitting the term structure using a flexible, smooth parametric function[2]. They demonstrate that their proposed model is capable of capturing many of the typically observed shapes that the yield curve adopts over time. Although this model was in essence designed to be a static model which does not account for the intertemporal evolution of the term structure, Diebold and Li (2006) show that the coefficients can be interpreted as three latent dynamic factors. Thus, the time-varying parameters $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ are referred to as latent dynamic factors or factor realizations and they have an economic interpretation in the term structure. $\{F_{1t}, F_{2t}, F_{3t}\}$ are termed factor loadings and are the weight functions of the dynamic coefficients. The complete parameter set for the NS specification is $\theta_t = \{\beta_{1t}, \beta_{2t}, \beta_{3t}, \lambda_t\}$. The λ parameter is the shape parameter and governs the exponential decay rate; in other words, it determines the location of the maximum or minimum curvature component.

---

[1] A Laguerre function consists of a polynomial times an exponential decay term.

[2] Nelson and Siegel (1987) show that, empirically, the NS model fits the data well, as shown by Nelson and Siegel (1987), and it performs relatively well in out-of-sample forecasting analysis (see, among others, Diebold and Li, 2006, and De Pooter *et al.*, 2007). The functional form of the yield curve proposed by Nelson and Siegel (1987) ensures a smooth curve rather than the saw-toothed one of the spline method, but it is still sufficiently flexible to fit the generally observed shapes of yield curves (Sundaram and Das, 2010). However, from a theoretical viewpoint the NS yield curve model is not necessarily arbitrage-free (e.g., see Bjork and Christensen, 1999, Coroneo *et al.*, 2011) and does not belong to the class of affine yield curve models (e.g., see Diebold *et al.*, 2004). In this line of reasoning, Coroneo *et al.* (2011) study the extent to which the NS curves accord with the no-arbitrage condition. They find that despite the lack of a theoretical arbitrage-free derivation, the NS model produces dynamically consistent curves. Nevertheless, our main aim in this paper is not to analyse the theoretical properties of the NS model but to study the triangle shape parameter multicollinearity-estimation procedure.

The loading on β_{1t} is F_{1t} and is unity. It can thus be viewed as a long-term factor and the dynamic factor β_{1t} can be interpreted as the level factor. It holds that $y_t(\infty)=\beta_{1t} \geq 0$. The loading on β_{2t} is F_{2t}, which is a decreasing function that starts at 1 for $\tau=0$ and then decreases monotonically and quickly to 0. It may thus be viewed as a short-term factor. Note that $y_t(0)=\beta_{1t}+\beta_{2t}>0$. The dynamic factor β_{2t} is related to the slope of the yield curve and is measured as: (i) the difference $y_t(\infty)-y_t(0)=-\beta_{2t}$, or (ii) the difference between the ten-year yield and the three-month yield. The loading on β_{3t} is F_{3t}, which starts at 0, then increases and, finally, decays to 0. It may thus be viewed as a medium-term factor. The dynamic factor β_{3t} is closely related to the yield curve curvature. Specifically, it is measured as: $2y_t(2\,year)-y_t(3\,year)-y_t(10\,year)$.

3. Estimating the Nelson-Siegel model

The traditional procedures used to estimate the NS model parameters can be summarized as follows: i) minimizing the sum of squared errors (SSE) using OLS over a grid of pre-specified values of λs (Nelson and Siegel, 1987); ii) minimizing SSE using linear regression conditional on a chosen fixed shape parameter λ (Diebold and Li, 2006; de Pooter, 2007; and Fabozzi *et al*., 2005) and iii) using nonlinear optimization techniques (Cairns and Pritchard, 2001). Annaert *et al*. (2013) propose combining a grid search to determine the value of the optimal shape parameter with a ridge regression in order to solve some of the estimation problems resulting from the traditional linear and nonlinear estimation of the model. More specifically, we address the nature of the multicollinearity problem in the following sections using both real and simulated data and extending the work of these authors by analysing the statistical distribution of the estimated model parameters. In the estimation analysis carried out in the empirical section we compare the performance of the OLS regression after having set a value for the nonlinear estimation and the grid search-ridge regression approach.

3.1. Linear regression with a fixed value of the shape parameter

The most widely used approach in the literature to estimate the NS model is fixing the value for the shape parameter, λ, so that the remaining parameters can be estimated by OLS. In this method, the objective function is:

$$\min_{\theta} \sum_{i=1}^{N} (f(\theta_i,\tau_i)-y_i)^2 \tag{4}$$

where $f(\theta_i,\tau_i)$ is the estimated yield in (2) and y_i is the observed yield. N is the length of the maturities vector. The parameter set is $\theta_t=\{\beta_{1t},\beta_{2t},\beta_{3t}\}$ with the shape parameter having been previously fixed so that the initial nonlinear model is transformed into a linear one. This estimation method is the most popular among market practitioners due to its simplicity.

Diebold and Li (2006), using monthly data, find the value of the shape parameter that maximizes the curvature component to be $\lambda = 0.0609$. Fabbozi *et al*. (2005) set it at 3 as the result of a grid search for the whole dataset. In this article, we have fixed λ at 0.0609 as in Diebold and Li (2006). The main advantages of this estimation approach are that it does not require starting values for the estimates and that it provides global optimal estimators. It does, however, suffer from the limitations mentioned in Annaert *et al*. (2013) and outlined above. In addition, Annaert et al. (2013) show that the fixed shape parameter might not be robust to different maturity vectors. Indeed, the optimal shape parameter may also vary over time. In this article, we compare the ridge regression with both traditional linear and nonlinear estimation, which are used as benchmarks.

3.2. Nonlinear regression

There are a few studies which estimate the NS model using nonlinear optimization (e.g., see Cairns and Pritchard, 2001), all of which produce very unstable estimates of the model. As outlined above, in the next section we examine the problems of using nonlinear procedures to estimate the NS model using both real and simulated data. For this purpose, we use the MATLAB *lsqcurvefit* function, with the target function for a specific day being (4).

Note that the estimated value of the model parameters can change in each period. The main problem with this method is that the optimization problem is not convex and has multiple local optima, as well as being very sensitive to the initial values. Accordingly, the nonlinear standard methods that are readily available in statistical packages are not appropriate for estimating the NS model.

3.3. Grid search with ridge regression

Due to the limitations of the two estimation methods set out above, we follow Annaert *et al.* (2013) and their ridge regression approach by combining the grid search of the shape parameter with the OLS regression to "free" the shape parameter. These authors re-estimate the remaining parameters for the optimal λ obtained in the grid search for the days when the degree of multicollinearity[3] among factor loadings is too high. In order to do that, we have to define the collinearity measure we test in the ridge regres-

[3] We define multicollinearity through the concept of orthogonality. When the regressors are orthogonal or uncorrelated, all the eigenvalues of the design matrix are equal to one and the design matrix is of full rank. If at least one eigenvalue is different from one, especially when equal to zero or near zero, then no orthogonality exists, meaning that multicollinearity is present. (More details in Vinod and Ullah, 1981).

sion estimations. We use the condition number, which is based on the eigenvalues of the NS regressors and which, following Annaert *et al.* (2013), is defined as:

$$\kappa(\mathbf{X})=\frac{eig_{\max}}{eig_{\min}} \tag{5}$$

where $\mathbf{X}$ is the matrix of the regressors in (2) defined as a general regression model ($\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$) and eigmax and eigmin are the maximum and minimum eigenvalues respectively of the regressors matrix $\mathbf{X}$. If $\mathbf{X}$, which includes F_{1t}, F_{2t}, F_{3t}, presents multicollinearity, that is to say the regressors are correlated, then the condition number will be different from 1. The higher the difference between the two eigenvalues, the higher the degree of multicollinearity. In this article, we follow the suggestion of Belsley (1991) and use a condition number of 10, equivalent to a correlation coefficient above 0.8, as a measure of the degree of multicollinearity. Then, if we detect collinearity, we implement ridge regression to avoid the resulting instability in the OLS estimates. This re-estimation results in a trade-off between the reduction in the variance and the increase in the bias of the estimators; according to Kutner *et al.* (2004), estimators with small variance are preferred to unbiased ones.

The ridge regression estimates vector is as follows:

$$\hat{\boldsymbol{\beta}}=[\mathbf{X'X}+k\mathbf{I}]^{-1}\mathbf{X'y} \tag{6}$$

where k is the ridge constant (small and positive), which can be estimated using an iterative procedure searching for the lowest positive number that makes the new condition number fall below the chosen threshold.

According to Annaert *et al.* (2013), this process can be implemented as follows:

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 that finds the lowest positive number, k, which makes the recomputed condition number fall below the threshold.
4. By adding a small bias, the correlation between the regressors will decrease and so will the condition number.

In short, this method tries to correct the problems resulting from multicollinearity between the factor loadings in the estimation of the NS model and also allows the shape parameter to move freely over time.

4. Empirical analysis

In this section, we estimate the NS model with a database consisting of time series of cross sections, using the three estimation methods described in Section 3. We first present the database, then we address the issue of multicollinearity and check for its presence when using the benchmark OLS method, in order to justify the use of ridge regression.

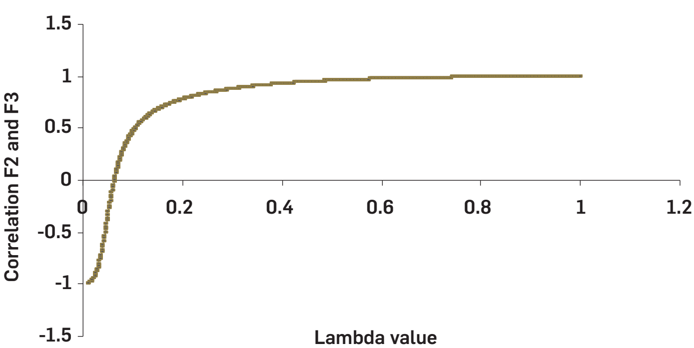
4.1. Interest rate data base: Initial analysis

We use the same database as in Diebold and Li (2006)[4], which contains the end-of-month price quotes (bid-ask average) for U.S. Treasuries from January 1985 through December 2000, taken from the CRSP (Center for Research in Security Prices) government bond files. Maturities are fixed at months τ=[3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120].

4.2. The collinearity problem

In this subsection, we study the correlation between the factor loadings of the NS model for the same database as in Diebold and Li (2006), as well as their evolution over time, in order to illustrate the collinearity problem in the OLS estimation procedure with a fixed λ value. Empirically, the three factors have been found to be only mildly correlated. This result is also corroborated by studies that build on principal component analysis (PCA) or assume zero correlations in factor analysis, yet still arrive at these factors, such as that by Litterman and Scheinkman (1991). When estimating the NS model, for many values of the λ-parameter the correlation between the second and the third loading is high (as is the condition number), thus the attribution of a particular yield curve shape to the specific factor becomes difficult.

Figure 1. **Correlation between factor loadings for $0<\lambda<1$ in the OLS estimation of the NS model**



[4] This database is available at http://www.ssc.upenn.edu/~fdiebold/YieldCurve.html

Figure 1 shows the correlation between the factor loadings for different values of λ ranging from 0 to 1 using the same maturities as in the Diebold and Li (2006) database. We observe that the correlation coefficient is –1 for $\lambda = 0$, and that it increases abruptly to 1 as λ grows (in fact, it practically reaches unity at $\lambda=0.4$).

In Diebold and Li (2006) the λ-parameter is fixed at 0.0609 (for a monthly database). For each cross-section of yields, they run an OLS regression, thus obtaining a time series of β-values. They then model the β-values as AR(1)-processes, and use these specifications to predict future β-values and hence yield curves. According to our results, their λ value is well-chosen, as it indicates only a weak positive correlation between the factor loadings. To show the consequences of high correlation, we replicate some of the results of Diebold and Li (2006) using a critical lambda value, such as $\lambda = 0.9$ (high correlation between the factor loadings).

Table 1 depicts a summary of the main descriptive statistics of the factors of the NS model for the OLS estimation procedure conditional on a shape parameter ranging between 0.0609 and 0.9. The estimated parameters become more unstable and much more volatile as the degree of multicollinearity increases. In short, the potential multicollinearity problem in the OLS estimation depends on the value of λ, so fixing its value is not an optimal solution. At any rate, if we aim to meaningfully estimate the parameters of the NS model, we need to restrict the λ values to ranges where practical identification is still possible: in this case between 0.06 and 0.1. Another more suitable solution to reduce the degree of multicollinearity and allow the λ parameter to be free is to use the ridge regression presented in subsection 3.3. The estimation results (and their comparison) as well as the forecasting analysis are studied in subsections 4.3 and 4.4, respectively.

Table 1. **Estimates of the NS model parameters. Diebold and Li (2006) database. OLS estimation with fixed shape parameter: λ=0.0609 and λ=0.9**

Factor	Mean	Std dev.	Minimum	Maximum
		OLS with λ = 0.0609 (Diebold and Li, 2006)		
β_1	7.5792	1.5242	4.4267	12.0887
β_2	-2.0983	1.6083	-5.6155	0.9190
β_3	-0.1623	1.6873	-5.2506	4.2327
		OLS with λ = 0.9 (Diebold and Li, 2006)		
β_1	7.1136	1.4415	4.3669	11.6095
β_2	19.3637	17.296	-11.4131	59.9616
β_3	-29.2828	24.7262	-86.8997	13.6209

4.3. Estimation results

Table 2 lists the main descriptive statistics of the estimated parameters resulting from the estimation methods detailed above. It can be observed that the nonlinear estimation produces the estimates of the parameters with the largest standard deviation, indicating a high degree of instability. The estimated parameters for the other models are more realistic and make more economic and statistical sense. More specifically, the estimates resulting from the ridge regression estimation are more stable, and have lower volatility than those provided by the linear regression after having fit the shape parameter λ (the procedure used in Diebold and Li, 2006). In addition, as mentioned above, the ridge regression allows the shape parameter to vary over time, which is more consistent with reality. Furthermore, this estimation approach prevents numerical problems in the optimization process due to multicollinearity. In sum, the performance in estimating the parameters of the NS model improves considerably when combining the grid search of the shape parameter with the ridge regression approach.

In the next subsection, we analyse the in-sample and out-of-sample forecasting ability of these three alternative estimation procedures in order to check whether the ridge regression approach outperforms the other two traditional estimation procedures.

Table 2. **Estimates of the NS model parameters with three alternative estimation approaches. Diebold and Li (2006) database.**

Factor	Mean	Std dev.	Minimum	Maximum
		Non-linear regression		
β_1	2.879	114.8458	-1472.8000	562.4165
β_2	3.3549	295.2528	-2610.9000	2706.2000
β_3	4.8221	296.3553	-2704.3000	2551.0000
λ	0.1206	0.5112	-0.2113	5.7932
		OLS with λ = 0.0609 (Diebold and Li, 2006)		
β_1	7.5792	1.5242	4.4267	12.0887
β_2	-2.0983	1.6083	-5.6155	0.919
β_3	-0.1623	1.6873	-5.2506	4.2327
		Lambda Grid search and ridge regression		
β_1	5.0917	1.2064	2.5091	8.5853
β_2	1.0411	0.8449	-0.4338	3.6184
β_3	0.9513	0.3625	0.2189	2.1122
λ	0.1099	0.1232	0.0101	0.596

4.4. Forecasting analysis

In order to compare the forecasting performance for the analysed estimation methods we use the Mean Absolute Prediction Error (MAPE). For every month in our database we estimate the NS model using the three competing procedures. Then, we employ the estimated term structures to forecast the rates used in the estimation (in-sample forecasting) and the contemporaneous yields (out-of-sample forecasting) using a rolling window procedure. The estimation procedure with the lowest MAPE will be the method with the best forecasting performance.

First, we focus on the in-sample forecasting analysis by computing the MAPE for the three estimation procedures considered. According to the results in Annaert *et al.* (2013), the linear regression with a fixed shape parameter outperforms the other two competing methods for all maturities (see Table 3).

Table 3. **In-sample MAPE. Estimation of the NS model with alternative estimation methods using the Diebold and Li (2006) database**

Maturity	OLS D-L	Nonlinear	Grid-Ridge
3	0.0711	0.2268	0.6640
6	0.0301	0.1855	0.5060
9	0.0533	0.1791	0.3886
12	0.0667	0.1671	0.3174
15	0.0653	0.1536	0.3172
18	0.0548	0.1285	0.3726
21	0.0369	0.1085	0.4262
24	0.0414	0.1172	0.4616
30	0.0295	0.0877	0.6818
36	0.0515	0.0926	0.8405
48	0.0624	0.0816	1.1536
60	0.0716	0.0919	1.3227
72	0.0649	0.0863	1.5463
84	0.0471	0.0827	1.6364
96	0.0456	0.0862	1.7648
108	0.0533	0.0895	1.8466
120	0.0616	0.1064	1.8610

Second, in order to investigate the ability of the estimation procedures to forecast the long and short term of the term structure, we focus on the out-of-sample performance test for a forecast horizon of one month. We compare the predicted rates for every maturity to the actual interest rate using the MAPE criterion. The results are presented in Table 4. In the out-of-sample forecast, the grid search with conditional ridge regression proposed by Annaert *et al.* (2013) outperforms the other models for maturities higher than one year. Therefore, the ridge regression implementation can also solve some estimation problems related to the collinearity between the factor loadings in the important out-of-sample forecasting framework.

Table 4. **Out-of-sample MAPE. Estimation of the NS model with alternative estimation methods using the Diebold and Li (2006) database**

Maturity	OLS D-L	Nonlinear	Grid-Ridge
3	1.7696	1.5821	1.8687
6	1.8219	1.6785	1.8310
9	1.8314	1.7213	1.7637
12	1.8441	1.7575	1.6864
15	1.8417	1.7714	1.6099
18	1.8623	1.8032	1.5714
21	1.8878	1.8385	1.5394
24	1.9120	1.8709	1.5135
30	1.9151	1.8864	1.4403
36	1.9327	1.9102	1.3862
48	1.9695	1.9634	1.3166
60	2.0348	2.0451	1.3049
72	2.0240	2.0476	1.2534
84	2.0296	2.0656	1.2367
96	2.0409	2.0884	1.2038
108	2.0786	2.1357	1.1869
120	2.1492	2.2131	1.1769

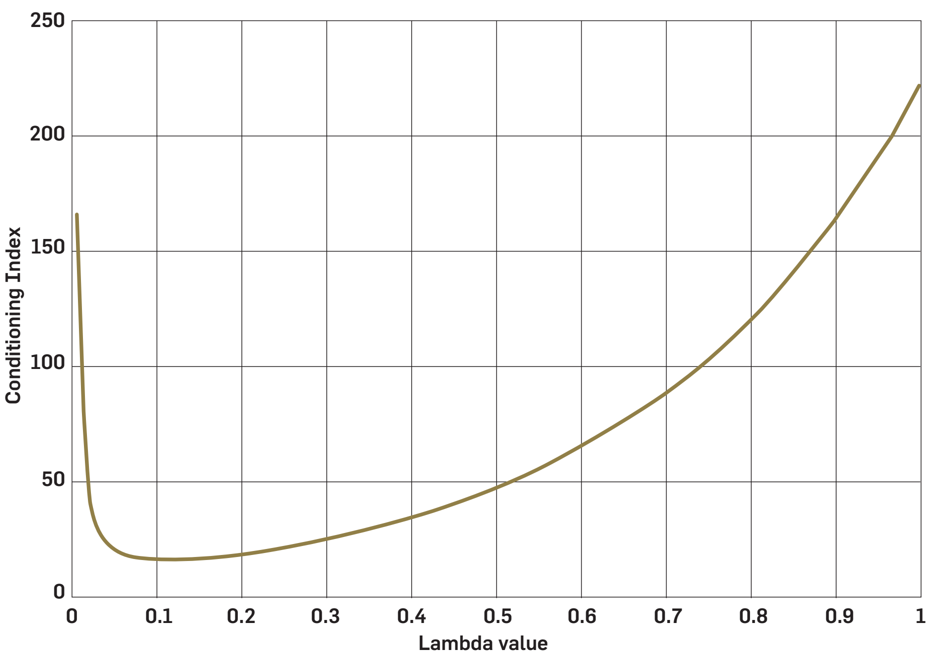
5. Simulation analysis

In this section, we carry out a Monte Carlo simulation exercise for the yield curve, based on the database in Diebold and Li (2006). The main aim is to analyse the deviations of the simulated parameters from the actual ones, thus comparing the alternative estimation procedures in the presence of multicollinearity. We generate 1000 errors for the model (2) using normal distributions with mean 0 and a small standard deviation. In the first stage, the error variance-covariance matrix will be diagonal as in Diebold *et al*. (2006) and Diebold and Rudebusch (2013). We generate a simulated yield curve database with these error distributions in order to obtain the statistical distribution of the estimated parameters and identify multicollinearity. In the second stage, we use the Cholesky decomposition to obtain the error variance-covariance matrix of the observed errors from the real dataset.

When using the diagonal error variance-covariance matrix, we first take a λ value which produces collinearity problems and compare the estimates resulting from the Diebold and Li (2006) approach with a fixed λ value, the nonlinear estimation and the grid search with conditional ridge regression. Taking the maturity vector $\boldsymbol{\tau}$ of the database used in the previous section and considering a grid of lambda values $\lambda \in (0,1)$ varying by 0.001, we analyse the behaviour of the conditioning index (CI) (calculated as the square root of the condition number defined in (5)) for the loading matrix $\mathbf{X}$ depending on the value of λ. In Figure 2, we show the evolution of this index conditional on the value of λ. From $\lambda = 0.05$, the higher the value of λ, the higher the value of the index. However, two critical areas can be identified: one for values very close to zero and another for values close to one.

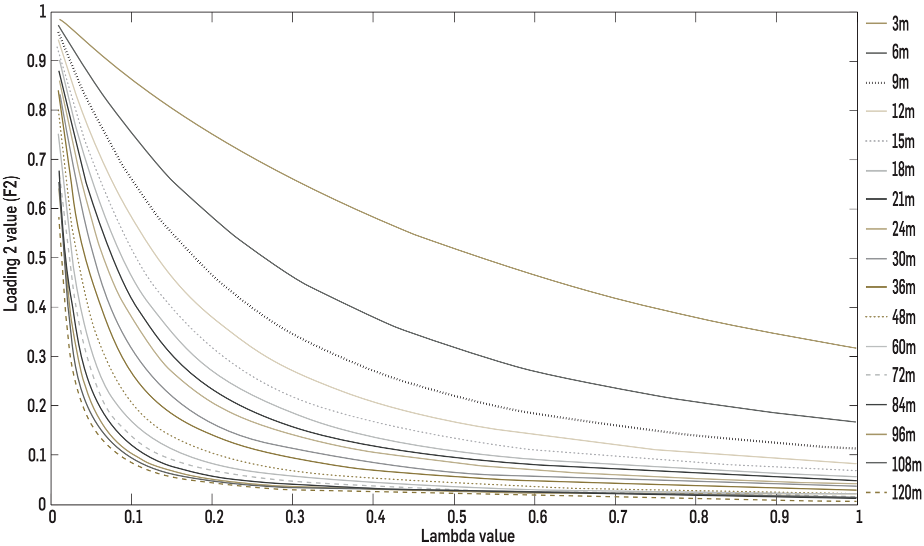
Figures 3 and 4 show how the factor loadings 2 and 3, respectively, evolve with the value of the shape parameter λ. In both Figures 3 and 4, it can be seen that the higher the value of λ, the higher the value of the factor loadings for long maturities. Therefore, the values of the factor loadings are conditional on the value of lambda, with the impact depending on the time to maturity.

Figure 2. **Conditioning index for a grid of λ values**

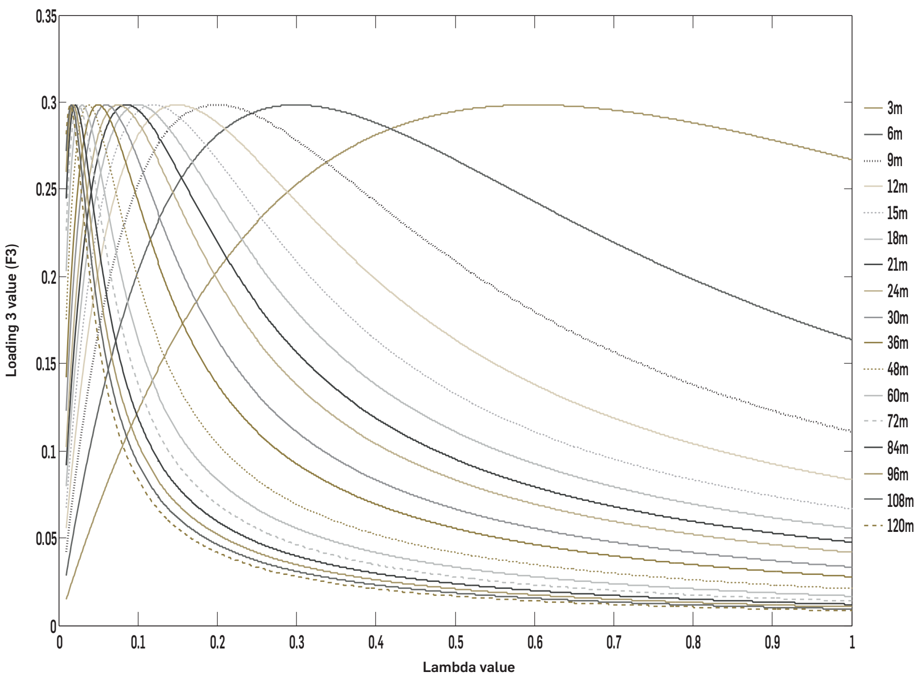


On multicollinearity and the value of the shape parameter in the term structure Nelson-Siegel model. *León, A., Rubia, A. and Sanchis-Marco, L.*

■ Figure 3. **Evolution of the loading factor 2 for different values of λ**



■ Figure 4. **Evolution of the loading factor 3 for different values of λ**



In order to examine the collinearity issue, we select a ‘conflictive’ value of λ resulting in a high conditioning index, for example $\lambda = 0.01$, with a CI value of 165.41. The data we use are simulated data from the errors generated with normal distributions using diagonal and nondiagonal variance-covariance. Table 4 of Diebold, *et al.* (2006) reports the standard deviations of the errors of the yield estimates, which we use to set the diagonal values of the variance-covariance matrix in the first stage. The simulation procedure entails the following steps: i) We generate 1000 random errors normally distributed for the 17 maturities considered; ii) The errors are computed as the product of the random errors and the diagonal variance-covariance matrix of the abovementioned errors; iii) The new simulation data are the sum of the NS-estimated parameters using the corresponding procedure (OLS with a fixed value for the shape parameter, nonlinear regression and ridge regression) and the simulated errors from step ii).

Table 5 lists the descriptive statistics for the estimated parameters of the NS model for the simulated data and the three competing estimation methods considered (for the linear regression, the shape parameter was fixed at $\lambda = 0.01$). As can be noticed, the best estimates are those produced by the grid search with conditional ridge regression. Whereas the estimates produced by the linear and (especially) nonlinear regression procedures are affected by multicollinearity, the ridge regression approach remedies this problem and stabilizes the estimates.

Table 5. **Descriptive statistics. Estimated parameters of the NS model using the simulated database (errors set in the diagonal V-C matrix taken from Diebold *et al.*, 2006)**

Factor	Mean	Std dev.	Minimum	Maximum
		Non-linear regression		
β_1	10.4205	3.9541	-6.1644	11.4105
β_2	-433.4383	1779.0000	-18256.0000	6283.2000
β_3	397.6026	1775.2000	-6334.6000	18208.0000
λ	3.9541	1.9397	0.0100	47.7567
		OLS with $\lambda = 0.01$ (Diebold and Li, 2006)		
β_1	-4.2450	1.1370	-8.2313	-0.8624
β_2	12.3624	1.1071	9.0685	16.2439
β_3	28.0432	1.6048	23.2707	33.6687
		Grid search and Ridge regression		
β_1	8.3023	0.0002	8.3017	8.3031
β_2	0.8102	0.0004	0.8084	0.8117
β_3	1.2307	0.0002	1.2301	1.2310
λ	0.1010	0.0000	0.1010	0.1010

Table 6 shows the descriptive statistics of the estimates obtained when simulating with the real covariance matrix of the estimated errors resulting from the NS model estimation using the Diebold and Li (2006) database. As can be observed, the results are very similar to those listed in Table 5. As outlined above, we next study the behaviour of the estimated parameters of the NS model, analysing the evolution of the difference between the estimates obtained by simulation and the real values, depending on the degree of multicollinearity between the factor loadings and the value of the shape parameter in the OLS estimation procedure with a fixed value of λ.

Table 6. Descriptive statistics. Estimated parameters of the NS model using the simulated database (errors set in the nondiagonal V-C matrix taken from Diebold *et al.*, 2006)

Factor	Mean	Std dev.	Minimum	Maximum
		Non-linear regression		
β_1	5.0189	3.7776	-5.9209	10.6331
β_2	-1310.4000	3125.6000	-20563.0000	486.2568
β_3	1312.4000	3108.9000	-525.1691	20524.0000
λ	0.9297	1.7272	0.0101	5.8641
		OLS with λ = 0.01 (Diebold and Li, 2006)		
β_1	-6.2669	1.1370	-10.2532	-2.8843
β_2	14.3307	1.1071	11.0368	18.2122
β_3	30.8988	1.6048	26.1263	36.5242
		Grid search and Ridge regression		
β_1	5.7115	0.3283	5.5742	8.6194
β_2	3.5979	0.3303	0.4356	3.7357
β_3	1.4672	0.2014	0.7574	2.0210
λ	0.0121	0.0069	0.0101	0.1717

More specifically, a value of 0.01 is set for λ in the OLS estimates of the NS model with the real database. Table 7 presents the descriptive statistics for these estimates, to be compared with those obtained with the other two competing models (Tables 5 and 6). A substantial departure of the estimates from the true values can be observed for the OLS method, and some of these deviations do not make economic sense; this is the case with the β_1 (long-term yield) negative values. We also study another representative value of λ (0.0609), the value used in Diebold and Li (2006). In this case, more stable estimates are obtained and, from the comparison of Table 2 and Table 8 (which shows the descriptive statistics for the estimates obtained with the simulated data and a nondiagonal covariance matrix for the errors, using the OLS method with λ=0.0609), it can be deduced that the departures from the actual parameter values become smaller. Therefore, multicollinearity results in a substantial increase in the departure of the estimates obtained by simulation from the real values; as such, the sensitivity of the estimates to the value of λ in the NS model can be clearly observed. On the other hand, if we compare the results for the alternative methods presented in Tables 5 and 6 (simulated data) with those listed in Table 2 (actual database) it can be seen that it is the nonlinear that shows the greatest difference. Conversely, the grid search combined with ridge regression strategy is the method that generates the smallest differences; in addition, the estimates make economic sense. In short, the use of ridge regression remedies the adverse effect of multicollinearity in the estimation of the NS model parameters and, in addition, it allows the shape parameter to range freely over time.

Table 7. Descriptive statistics. Estimated parameters of the NS model with the Diebold and Li (2006) database using OLS estimation with a fixed shape parameter

OLS with λ = 0.01 (Diebold and Li, 2006)				
Factor	**Mean**	**Std dev.**	**Minimum**	**Maximum**
β_1	1.3015	4.6970	-9.9989	14.6009
β_2	4.3186	4.8297	-7.4617	15.7271
β_3	12.1117	9.2219	-12.6888	34.0332

Table 8. Descriptive statistics. Estimated parameters of the NS model obtained with the simulated database (errors set in the nondiagonal V-C matrix taken from Diebold *et al.*, 2006) using OLS estimation with a fixed shape parameter

OLS with λ = 0.0609 (Diebold and Li, 2006)				
Factor	**Mean**	**Std dev.**	**Minimum**	**Maximum**
β_1	11.3751	0.0013	11.3715	11.3797
β_2	-3.6642	0.0013	-3.6702	-3.6595
β_3	1.0008	0.0053	0.9851	1.0160

6. Concluding remarks

This article extends the study by Annaert *et al.* (2013) with an in-depth analysis of the distribution of the estimates produced by the NS model conditional on the λ parameter. In a preliminary analysis, we study the out-of-sample forecasting performance of three different approaches — the ridge regression method, nonlinear optimization and OLS estimation — for the NS model. The empirical evidence indicates that the ridge regression approach produces the best out-of-sample forecasting performance when maturities range between one and ten years.

Furthermore, we analyse the multicollinearity issue in the NS model factor loadings, which explain the suitability of a time-varying λ parameter. To that end, we carry out a Monte Carlo simulation exercise, generating a yield database from the Diebold and Li (2006) estimation errors in the NS model. The main results of this simulation exercise are: i) when using the OLS method with external determination of the shape parameter, multicollinearity results in unstable estimates that do not make economic sense ; ii) a grid search with conditional ridge regression corrects the problems resulting from collinearity between the factor loadings of the NS model and, in addition, it allows for a free shape parameter; iii) the nonlinear procedure produces unstable parameter estimates due to optimization problems.

The results are interesting for policy-makers, forecasters and practitioners, enabling them to draw ever more precise, stable and accurate determinations from the yield curve information using the NS model. According to our findings, the estimation procedures used in the financial sector can be substantially improved by remedying the collinearity problem and also allowing a variable shape parameter.

This paper can be extended using the Svensson model (1994), characterized by four model parameters and two shape parameters, and other specifications such as the NS model estimated in studies by De Potter (2007), Bliss (1997), Bjork and Christensen (1999) and Diebold, *et al.* (2008). Further interesting research would be to allow for a time-varying setting for the dynamics of the beta parameters. This would generate a dynamic model in line with Diebold, *et al.* (2006), who propose a new framework of modelling the yield curve under a state-space system using the Kalman filter. This new framework enables the VaR to be forecasted for portfolios of bond returns, thus enabling an evaluation of the performance of the extended two-step approach in Diebold and Li (2006) (comparing their results with those driven by the data generating process, the state-space system).

Finally, we can assume that the true data generating process is driven by the dynamic NS model with multivariate stochastic volatility for the errors of the transition equation, following Koopman (2010) and Hautsch and Yang (2010). Hence, the idea is to estimate the extended two-step Diebold and Li (2006) model and a multivariate GARCH process for the system 3x1 vector time series with components $\{\beta_{k,t}\}_{t=1}^{T}$. For instance, the popular parsimonious DCC model proposed by Engle (2002) could be estimated. Other possible alternatives for modelling multivariate GARCH can be seen in Engle (2009).

References

- Abeling, P. (2013). The optimal Nelson-Siegel model within the Solvency II framework, White paper series, RiskQuest.
- Annaert, J., Claes, A., De Ceuster, M. and Zhang, H. (2013). Estimating the spot rate curve using the Nelson-Siegel model: a ridge regression approach, *International Review of Economics and Finance*, **27,** pp. 482-496
- BIS (2005). Zero-Coupon Yield Curves: Technical Documentation, Bank for International Settlements, Basel.
- Björk, T. and Christensen, B. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves, *Mathematical Finance,* **9**, pp. 323-348.
- Bliss, R.R. (1997). Testing Term Structure Estimation Methods, *Advances in Futures and Options Research,* **9**, pp. 197-231.
- Cairns, A.J.G. and Pritchard, D.J. (2001). Stability of Descriptive Models for the Term Structure of Interest Rates with Applications to German Market Data, *British Actuarial Journal*, **7**, pp. 467-507.
- Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2011). How arbitrage-free is the Nelson-Siegel model?, *Journal of Empirical Finance*, **18**(3), pp. 393-407.
- Cox, J., Ingersoll, J.E. and Ross, S.A. (1985). A Theory of the Term Structure of Interest Rates, *Econometrica*, **53**, pp. 385-407.
- De Pooter, M. (2007). Examining the Nelson-Siegel class of term structure models, Tinberger Institute Discussion Paper, TI 2007-043/4.
- De Pooter, M., Ravazzolo, F. and van Dijk, D. (2007). Predicting the Term Structure of Interest Rates: Incorporating Parameter Uncertainty, Model Uncertainty and Macroeconomic Information, Tinbergen Institute Discussion Paper TI 2007-028/4.
- Diebold, F.X., Ji, L., and Li, C. (2004). A Three-Factor Yield Curve Model: Non-Affine Structure, Systematic Risk Sources, and Generalized Duration. In L.R. Klein (ed.), *Long-Run Growth and Short-Run Stabilization: Essays in Memory of Albert Ando*. Edward Elgar, Cheltenham, U.K., pp. 240-274.
- Diebold, F.X. and Rudebusch, G.D. (2013). *Yield curve modelling and forecasting: the dynamic Nelson-Siegel approach*, Princeton University Press, Princeton.
- Diebold, F.X., Rudebusch, G.D. and Aruoba, B. (2006). The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach, *Journal of Econometrics*, **131**, pp. 309-338.
- Diebold, F.X. and Li, C. (2006). Forecasting the term structure of Government bond yields. *Journal of Econometrics*, **130**(2), pp. 337-364.
- Diebold, F.X., Li, C. and Yue, V.Z. (2008). Global yield curve dynamics and interactions: a generalized Nelson-Siegel approach, *Journal of Econometrics*, **146**, pp. 351-363.
- Engle, R.F. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics*, **20**(3), pp. 339-50.

■ Engle, R.F. (2009). *Anticipating correlations*, Princeton University Press, Princeton.

■ European Central Bank (2008). The new Euro area yield curves, Monthly bulletin, February, pp. 95-103.

■ Fabozzi, F.J., Martellini, L. and Priaulet, P. (2005). Predictability in the shape of the term structure of interest rates, *Journal of Fixed Income*, **15**(1), pp. 40-53

■ Ferstl, R. and Weissensteiner, A. (2011). Asset-liability management under investment opportunities, *Journal of Banking and Finance*, **35**(1), pp. 182-192.

■ Goukasian, L. and Cialenco, I. (2006). The reaction of term structure of interest rates to monetary policy actions, *Journal of Fixed Income*, **16**(2), pp. 76-91.

■ Hautsch, N. and Yang, F. (2010). Bayesian inference in a stochastic volatility Nelson-Siegel model, SFB 649 Discussion Paper 2010-004, Humboldt University, Berlin.

■ Hull, J. and White, A. (1990). Pricing Interest Rate Derivative Securities, *Review of Financial Studies*, **3**, pp. 573-92.

■ Jakas, V, (2011). Theory and Empirics of an affine Term Structure. Model Applied to European Data, *AESTIMATIO, The IEB International Journal of Finance*, **2**, pp. 2-19.

■ Koopman, J.S., Mallee, M.I.P. and 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 and Economic Statistics*, **28**(3), pp. 329-343.

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

■ Litterman, R. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns, *Journal of Fixed Income*, **1**(1), pp. 54-61.

■ Martellini, L. and Meyfredi, J.C. (2007). A copula approach to Value-at-Risk estimation for fixed-income portfolios, *Journal of Fixed Income*, **17**(1), pp. 5-15.

■ Molenaars, T.K. and Reinerink, N.H. and Hemminga, M.A. (2013). Forecasting the yield curve - Forecast performance of the dynamic Nelson-Siegel model from 1971 to 2008. Available at https://mpra.ub.uni-muenchen.de/61862/.

■ Molenaars, T.K., Reinerink, N.H. and Hemminga, M.A. (2015). Forecasting the yield curve: art or science? Available at https://mpra.ub.unimuenchen.de/61917/1/MPRA_paper_61917.pdf

■ Mumtaz, H. and Surico, P. (2009). Time-varying yield curve dynamics and monetary policy, *Journal of Applied Econometrics*, **24**(6), pp. 895-913.

■ Nelson, C. and Siegel, A.F. (1987). Parsimonious modelling of yield curves, *Journal of Business*, **60**, pp. 473-489.

■ Rezende, R.B. and Ferreira, M.S. (2011). Modeling and Forecasting the yield curve by an extended Nelson-Siegel class of models: a Quantile Auto-regression Approach, *Journal of Forecasting*, **32**(2), pp. 111-123.

■ Svensson, L.E.O. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994, NBER Working Paper Series, 4871.

■ Sundaram, R.K. and Das, S.R. (2010). *Derivatives: principles and practice*, McGraw-Hill Irwin, New York.

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

■ Vinod, H.D. and Ullah, A. (1981). *Recent Advances in Regression Models*, Marcel Dekker, New York.

■ Yu, W.C. and Salyards, D. (2009). Parsimonious Modeling and Forecasting of Corporate Yield Curve, *Journal of Forecasting*, **28**(1), pp. 73-88.

■ Yu, W.C. and Zivot, E. (2011). Forecasting the term structures of Treasury and corporate yields using dynamic Nelson-Siegel models, *International Journal of Forecasting*, **27**(2), pp. 579-591.