Extrapolation of Long-Term Rates and Their Volatilities with the Svensson Model

Jan Annaert [*,**] , Anouk G.P. Claes [***] , Marc J.K. De Ceuster [°,*,**] and Hairui Zhang [*]

February 12 2013

Abstract

The Nelson-Siegel and the Svensson models are widely used in practice for fitting the term structure of interest rates. In this paper, a conditional ridge regression based approach is proposed to extrapolate the long end of the yield curve. Judged by the fitting errors between the rates and the volatilities of the 30-year swap rate and its model-extrapolated counterpart, the method we suggest is more accurate to describe the long-term rates on the yield curve than traditional algorithms to estimate the Svensson model. The accuracy is not only reflected on interest rate level, but also on conditional volatility fit and forecast.

Keywords: Term Structure, Nelson-Siegel Model, Svensson Model, Ridge Regression

JEL Code: C13, E47

* Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerp, Belgium
** Antwerp Management School, Sint-Jacobsmarkt 9, 2000 Antwerp, Belgium
*** Saint-Louis University Brussels, part of the Louvain School of Management Group, Boulevard du Jardin Botanique 43, 1000 Brussels, Belgium

° Corresponding author: marc.deceuster@ua.ac.be

1. Introduction

Nelson and Siegel (1987) propose a parsimonious parametric specification to describe the shape of the term structure of interest rates. With only four parameters, the model is able to describe a whole family of empirically observed term structure shapes. Moreover, the additive terms in the model can be interpreted as a level, slope and curvature factor, commonly acknowledged through factor analysis (Litterman and Scheinkman, 1991). Svensson (1994) adds an additional curvature term to the Nelson-Siegel specification so that it can fit an even broader set of term structures. According to the Bank of International Settlements (2005), both models are heavily used by central banks. Finland and Italy were reported to use the Nelson-Siegel model; Germany, Norway, Spain and Switzerland resorted to the Svensson model; and Belgium, France and Sweden opted for both depending on their fit.

The econometric estimation of these models, however, copes with serious unreported or often loosely reported difficulties. Both models suffer from a high degree of nonlinearity as well as from potential multicollinearity. Not surprisingly, Barrett, Gosnell and Heuson (1995), Bolder and Stréliski (1999), Cairns and Pritchard (2001), Fabozzi, Martellini and Priaulet (2005), Gurkaynak, Sack and Wright (2006), de Pooter (2007) and Gilli, Grosse and Schumann (2010), all report numerical instabilities. Not only do parameter estimates turn out to be heavily dependent on starting values, but both models also produce time series of parameter estimates that are highly erratic over time. Linearizing the Nelson-Siegel model - by fixing the shape parameter at a level at which the factors are only moderately correlated - has become common practice (c.f. Diebold and Li, 2006 and Fabozzi, et al., 2005 for the Nelson-Siegel model). This approach, however, comes with a loss of flexibility in the shape parameter and might not necessarily be inspired by economic reasons. Annaert, Claes, De Ceuster and Zhang (2013) suggest a two-step procedure, which can be described as a conditional ridge regression, to cope with the reported problems for the Nelson-Siegel model. A grid search over the shape parameter is implemented using OLS. Conditional on the degree of multicollinearity implied by the choice of the 'optimal' shape parameter, the equation is re-estimated using ridge regression. This simple approach allows the shape parameter to be freely estimated and alleviates the instability in parameter estimation caused by multicollinearity. At least for the Nelson-Siegel model, they find a seriously improved fit, especially in the extrapolated long rates.

Despite the work of Ferenczi and Werner (2006), the estimation of the Svensson model has not received a lot of attention. Even Ferenczi and Werner (2006) do not address the multicollinearity problem in the Svensson model. As the Svensson model shares all the problems reported in implementing the Nelson-Siegel model, the conditional ridge regression approach can be thought of as a natural candidate to improve the estimated Svensson model coefficients and their stability. This paper assesses the estimation performance of the conditional ridge regression approach compared to other estimation procedures. We provide empirical evidence that within sample, a two-dimensional grid search has the lowest mean absolute errors. However, an estimation procedure that selects - based on their in-sample performance - either the Svensson model, using a two-dimensional grid search and ridge regression, or the Nelson-Siegel model with ridge regression, renders the most accurate out-of-sample performance for the long end of the yield curve.

Annaert et al. (2013) have shown the advantages of applying ridge regression on the Nelson-Siegel model, especially when extrapolating long-term rates. Here we take one step further to estimate the Svensson model with ridge regression.

Our attention is paid on extrapolation of long-term rates with the Nelson-Siegel-Svensson models. We do this for several reasons.

First of all, the long end of the yield curve is important to both monetary policy makers as well as companies whose main business deals with long-term financial products. For example, lowering short-term interest rates or maintaining a low level of short-term rates, a common practice of monetary policy makers nowadays thanks to the financial crisis, forces investors to borrow at short-term rates to invest in longer-term investments, driving down long-term rates and boosting short-term rates, leading to a yield curve that is flattened or with a negative slope. Conventional wisdom tells us that a negative slope is considered as an indicator for economic recession. In this case, understanding how long-term rates behave can be used as a gauge to measure the potency of monetary policy actions on the general economy.

As another example, mortgage loans can be viewed as options which are sensitive to long-term interest rate falling. When long-term interest rates go down, homeowners are likely to refinance their mortgage loans, increasing interest rate exposure of banks and other financial institutions who sell mortgage-related products.

Long-term interest rates are also important for life insurance companies and pension funds who sell long-term guarantee products. If a person purchases a life insurance policy at the age of 20, the policy would likely be associated with long-term rates up to 50 years or more. Besides, long-term rates are also important to the liability side of these companies as lower long-term rates will boost the present value of their liability, making them difficult to meet their solvency requirement.

Nevertheless, unlike short-term rates such as the Euro Overnight Index Average (EONIA) rates that are observable from the money market, extrapolation is necessary to obtain long-term rates. For example, with Euro swap rates we are only able to bootstrap spot rates up to 10 years. For rates with maturity longer than

10 years, we need to extrapolate. For this reason, it is important to evaluate the extrapolation power of a model.

Meanwhile, Díaz, Jareño and Navarro (2011) show that alternative yield curve estimation techniques also have a serious impact on the estimates of the term structure of volatilities. We show that the conditional ridge regression approach that we advance also provides a more accurate fit and one-day ahead forecasts for the long term interest rate volatilities.

The paper is organized as follows. In Section 2 we introduce the extended Nelson-Siegel model, also known as the Svensson model. Next, in Section 3, we present eight possible estimation procedures. The data are described in Section 4. Section 5 discusses our empirical results. Finally, we conclude.

2. The Svensson Model

The Svensson model extends the Nelson-Siegel model with an extra factor that represents an additional hump/trough. The extra term allows describing an even broader family of yield curves than the Nelson-Siegel model.

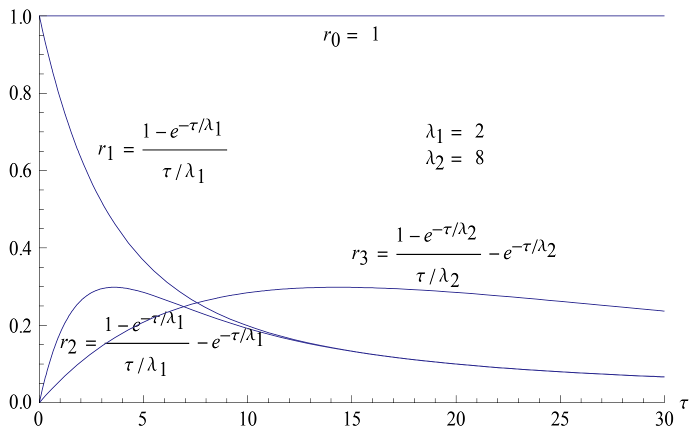
The spot rate function reads

$$r(\tau)=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{bmatrix}'\begin{bmatrix}1\\ \lambda_1\left(1-e^{-\tau/\lambda_1}\right)/\tau\\ \lambda_1\left(1-e^{-\tau/\lambda_1}\right)/\tau-e^{-\tau/\lambda_1}\\ \lambda_2\left(1-e^{-\tau/\lambda_2}\right)/\tau-e^{-\tau/\lambda_2}\end{bmatrix}\equiv\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{bmatrix}'\begin{bmatrix}r_0\\ r_1\\ r_2\\ r_3\end{bmatrix}, \tag{1}$$

where $r(\tau)$ is the continuous spot rate with time to maturity τ, $\beta_0,\beta_1,\beta_2,\beta_3,\lambda_1$ and λ_2 are the coefficients, with $\lambda_1>0,\lambda_2>0$. The first three components are exactly the same as those in the Nelson-Siegel model. The constant can be interpreted as a level component, the exponential models the slope of the curve and the Laguerre function (i.e. the product of an exponential and a polynomial function) allows for the presence of a hump/trough. The last component, $\beta_3 r_3$, adds a second hump/trough to the standard Nelson-Siegel model. As λ_1 and λ_2 determine the location of the two humps/troughs, they are also called shape parameters.

Figure 1 depicts the four building blocks of the Svensson model. The curves $r_0, r_1, r_2,$ and r_3 represent the level, slope, first and second curvature component of the spot rate curve.

Figure 1: Decomposition of the Svensson model with the shape parameters fixed at 2 and 8



Note: This figure shows the components of the Svensson model for the spot rate curve when the shape parameters are fixed at 2 and 8. The curves r_0, r_1, r_2 and r_3 *represent the level, slope and two curvature components of the spot rate curve.*

As both ends of the two curvature components are zero, the Svensson model has the same boundary conditions as the Nelson-Siegel model, which means that $r(0) = \beta_0 + \beta_1$ and $r(\infty) = \beta_0$. These boundary conditions imply that the long-term interest rate level is β_0, while the short-end of the yield curve is $\beta_0 + \beta_1$. As a result, the slope of a yield curve is given by $-\beta_1$. A negative (positive) β_1 means the term structure is upward (downward) sloping. The two shape parameters determine both the shape of the curvature components and the locations of humps/troughs of the term structure.

3. Estimation Procedures

3.1. Fitting the Term Structure of Interest Rates

The Svensson model has been estimated using various econometric techniques, including maximum likelihood estimation (MLE), nonlinear least square (NLS), amongst others. Bolder and Stréliski (1999) use both NLS and MLE. Ramponi and Lucca (2003) resort to NLS. Ferenczi and Werner (2006) perform a two-step optimization procedure in which the two shape parameters are estimated simultaneously and then the Svensson model is linearized to obtain the other estimates. Gurkaynak et al. (2006) use MLE to minimize the sum of squared errors between the estimated and the actual prices of Treasury securities, where the prices are weighted by the inverse of their durations.

Since it is just an extension of the Nelson-Siegel model, the Svensson model shares all the reported estimation problems. Furthermore, in the Svensson model there may not only be correlation between the slope and two curvature factors but also the curvature components themselves can be correlated. In this paper, we compare eight methods to estimate the yield curve and its conditional volatilities. In two subsections, we discuss procedures to estimate the Nelson-Siegel model (Section 3.1.1) and the Svensson model (Section 3.1.2).

3.1.1. Econometric Procedures for Estimating the Nelson-Siegel Model

Nelson and Siegel (1987) linearize their model by fixing the shape parameter λ_1 which allows them to estimate their model using ordinary least squares (OLS). For a pre-defined grid on λ_1, the OLS estimate with the lowest sum of squared errors is considered to be the best possible estimate. We refer to this procedure as a Grid Search (GS). The grid on λ_1 that we use, spans the interval (0, 10]. In order to speed up the estimation, the steps in the grid are being determined by MATLAB's FMINBND routine which is based on a golden section search with parabolic interpolation.

The degree of correlation between the regressors in the linearized Nelson-Siegel model, however, is extremely sensitive to the choice of λ_1. As documented by Annaert et al. (2013), re-estimation of the model using ridge regression conditional upon the presence of multicollinearity, improves the estimation of the (extrapolated) short and long rate significantly. Hence, based on the GS estimated λ_1, we measure the degree of multicollinearity using the condition number κ (kappa).

In a linear system $\mathbf{y} = \mathbf{b}'\mathbf{X}$ where each independent variable in $\mathbf{X}$ is standardized (centered and scaled), the condition number κ is given as:

$$\kappa(\mathbf{X}) = \frac{\mathbf{V}_{\max}}{\mathbf{V}_{\min}} \geq 1, \tag{2}$$

where $\mathbf{V}$ are the eigenvalues of $\mathbf{X}'\mathbf{X}$. If $\mathbf{X}$ is well-conditioned (i.e. the regressors are uncorrelated), then the condition number is one, which implies that the variance of $\mathbf{y}$ is explained equally by all the regressors. If correlation exists, then the condition number is no longer equal to 1 and the difference between the maximum and minimum eigenvalues grows as the collinearity increases. If the condition number is higher than the pre-defined threshold (e.g. 10 in Annaert et al., 2013), the model is re-estimated using ridge regression. [1,2] We refer to this method as a Grid Search augmented with ridge regression (GS-RR).[3]

3.1.2. Econometric Procedures for Estimating the Svensson Model

In analogy to the GS procedure for estimating the Nelson-Siegel model, the Svensson model can also be linearized by first fixing *both* shape parameters. Several procedures come to mind for fixing these parameters. Following Ferenczi and Werner (2006), we can linearize the Svensson model by performing an OLS-grid search on both λ_1 and λ_2 simultaneously. The parameter estimates resulting in the lowest sum of squared errors are deemed to be 'optimal'. We shall refer to this method as a two-dimensional grid search (2dimGS). Notice that in the two-dimensional grid λ_1 can be chosen to be equal to λ_2. In that case the Svensson model collapses to the Nelson-Siegel model. In our empirical work, the grid on the shape parameters used spans the interval (0, 10]. The steps in the grid are determined by MATLAB's LSQNONLIN routine.

Alternatively, however, one could estimate λ_1 first, by performing a GS estimation of the Nelson-Siegel model and then use this estimated λ_1 in the Svensson model to perform a second one-dimensional grid search on λ_2. This method we refer to as a double one-dimensional grid search (2×GS). Based on the same cut-off criterion on the condition-number as in GS-RR, we apply ridge regression whenever necessary in the 2×GS and the 2dimGS procedures. These procedures are labeled 2×GS-RR and 2dimGS-RR.

The Svensson model adds flexibility to the Nelson-Siegel model in describing the shape of the yield curve. Following the principle of parsimony, however, we prefer models with fewer parameters. The grid search based Svensson models always yields a lower sum of squared errors (SSE) than those obtained from the Nelson-Siegel model, because of the additional degree of freedom by the second curvature component. However, this need not be the case for the procedures where we use ridge regression. As the Svensson model may experience more multicollinearity problems (compared to the Nelson-Siegel model), more bias has to be added into the estimates in order to lower their variance. To determine whether the Svensson model's extra hump/trough is desirable, we finally compare the in-sample fitting errors produced by the

---

[1] Appendix 1 explains the implementation of the ridge regression.
[2] A threshold of 10 and 20 are both tested in this paper. The results with a threshold of 20 are very similar to those reported here. They are available upon request.
[3] Diebold and Li (2006) and Fabozzi, et al. (2005) fix the shape parameter λ_1 at 1.37 and 3 respectively so that no high degree of multicollinearity arises among the explanatory variables. As pointed out by Annaert et al. (2013), GS-RR outperforms a fixed shape parameter algorithm in terms of fitting accuracy. The results based on fixed shape parameter of 1.37 and 3 are available upon request.

ridge regression procedure based on the Nelson-Siegel model (GS-RR) with the ones based on the Svensson model (2×GS-RR or 2dimGS-RR). Whichever model produces the lowest sum of in-sample errors is favored. These selection procedures we refer to as 2×GS/GS-RR and 2dimGS/GS-RR. Table 1 summarizes the eight methods used in this paper.

Table 1 Summary of estimation methods

Name	Algorithm
GS	Grid search on the Nelson-Siegel specification
GS-RR	GS augmented with ridge regression
2×GS	1-dimensional grid search on Svensson with GS shape parameter
2dimGS	2-dimensional grid search on the Svensson specification
2×GS-RR	2×GS augmented with ridge regression
2dimGS-RR	2dimGS augmented with ridge regression
2×GS/GS-RR	2×GS-RR or GS-RR determined by the smaller in-sample sum of squared errors
2dimGS/GS-RR	2dimGS-RR or GS-RR determined by the smaller in-sample sum of squared errors

Note: This table summarizes the estimation methods that we use in this paper.

3.1.3. Evaluation of the Estimation Procedures

Using the procedures described in the previous subsections, we can fit a Svensson-type term structure model for the spot rate specification in Equation (1) for each day in our sample. Since we focus on extrapolation of long-term rates, in-sample fitting quality is not our main concern. We report the in-sample mean absolute error (MAE), but we put more focus on the 'out-of-sample' performance of the models. Following Annaert et al. (2013), we extrapolate from the fitted curves 30-year swap rates and compute the MAE vis-à-vis the contemporaneous 30-year ones. The model with the lowest MAE is considered to be the preferred model.

3.2. Fitting the Term Structure of Conditional Volatilities

To evaluate the usefulness of the proposed estimation procedures, we use the estimated term structures in two additional ways. We start by modeling the time series of the 30-year swap rate by fitting an exponential generalized autoregressive conditional heteroscedasticity (EGARCH) specification with student-*t* innovations. Díaz et al. (2011) compare several conditional volatility specifications to model the term structure of volatilities. They find that an EGARCH model with Gaussian innovations provides the lowest information loss in their dataset. We test GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) with both Gaussian and student-*t* innovations. Since the EGARCH(1,1) model with student-*t* innovations provided the best fit for all estimation methods in our paper, we choose this specification in this paper.

We proceed by fitting the same model through the time series of the extrapolated 30-year swap rates. By computing the MAE between the conditional volatility of the observed rates with the conditional volatility estimates on the extrapolated rates, we test the ability of the estimation procedures to capture the volatility dynamics at the long end of the yield curve. The model with the lowest MAE is considered as most accurate to describe conditional volatility.

Finally, we forecast one-day ahead conditional volatilities for the 30-year swap rate. We evaluate the one-day volatility forecasting ability of the estimation procedures. Again, the procedure with the lowest forecasting errors as measured by MAE is considered as the most appropriate one to fit the term structure of volatilities.

4. Data

We gathered the Euribor rates maturing from 1 week up to 12 months and Euro swap rates with maturities between 2 and 10 years. The 30-year Euro swap rates were also collected to assess the 'extrapolation' quality of the eight estimation procedures. The Euribor and Euro swap rates were retrieved from Thomson Reuters DataStream®. Our dataset spans the period from January 4, 1999 to May 24, 2011 and includes 3 174 days. Following Annaert et al. (2013), we use the smoothed bootstrap procedure to construct the spot rate curve.

Panel A in Table 2 presents the descriptive statistics of the time series of the (continuously compounded) spot rates. The panel shows that the volatility of the time series decreases from 1.31% for weekly rates to 0.79% for the 10-year spot rate. The average spot rate varies from 2.71% for the one-week rate, to 4.33% for a 10-year maturity. Autocorrelation is close to 1 for the rates of all maturities, indicating a large persistence. Panel B in Table 2 summarizes the descriptive statistics of the spread between 10-year and 1-week rates, between 5-year and 1-week rates, and between 10-year and 5-year rates. Generally speaking, the yield curve is upward sloping, with a larger spread between short- and mid-term yields. The low serial correlation with a lag of 255 days as well as the wide range of these spreads reflect the time-variation of the yield curve shape.

Table 2 Descriptive statistics of spot rates (in percentage)

Panel A							
Maturity	**Mean**	**Std. Dev.**	**Min.**	**Max.**	$\hat{\rho}(5)$	$\hat{\rho}(25)$	$\hat{\rho}(255)$
1 Week	2.71	1.31	0.34	5.17	0.995	0.978	0.426
1 Month	2.81	1.33	0.39	5.36	0.996	0.971	0.413
2 Months	2.87	1.33	0.51	5.31	0.997	0.977	0.402
3 Months	2.93	1.31	0.64	5.48	0.998	0.978	0.393
4 Months	2.96	1.29	0.75	5.50	0.998	0.979	0.392
5 Months	2.99	1.27	0.85	5.41	0.998	0.979	0.391
6 Months	3.01	1.25	0.96	5.43	0.998	0.979	0.388
7 Months	3.03	1.24	1.00	5.42	0.998	0.979	0.387
8 Months	3.05	1.23	1.04	5.44	0.998	0.979	0.385
9 Months	3.07	1.22	1.10	5.43	0.998	0.979	0.382
10 Months	3.09	1.21	1.13	5.44	0.998	0.979	0.380
11 Months	3.10	1.20	1.17	5.45	0.998	0.978	0.377
12 Months	3.12	1.19	1.22	5.45	0.998	0.978	0.373
2 Years	3.25	1.10	1.21	5.43	0.995	0.970	0.423
3 Years	3.46	1.01	1.35	5.51	0.994	0.966	0.431
4 Years	3.63	0.95	1.53	5.56	0.994	0.963	0.442
5 Years	3.78	0.90	1.71	5.61	0.993	0.962	0.456
6 Years	3.92	0.87	1.87	5.69	0.993	0.962	0.475
7 Years	4.04	0.84	2.02	5.75	0.993	0.962	0.495
8 Years	4.15	0.82	2.15	5.82	0.993	0.963	0.514
9 Years	4.25	0.81	2.26	5.89	0.992	0.963	0.529
10 Years	4.33	0.79	2.35	5.96	0.992	0.963	0.539
Panel B							
Spread	**Mean**	**Std. Dev.**	**Min.**	**Max.**	$\hat{\rho}(5)$	$\hat{\rho}(25)$	$\hat{\rho}(255)$
10y – 1w	1.62	0.86	-0.67	3.40	0.986	0.942	0.166
5y – 1w	1.07	0.69	-0.83	2.68	0.977	0.909	0.007
10y – 5y	0.55	0.26	-0.22	1.04	0.987	0.943	0.362

Note: Panel A reports summary statistics of bootstrapped spot rates. Panel B reports summary statistics of the spreads between 10-year and 1-week rates, 5-year and 1-week rates, and 10-year and 5-year rates. Spot rates are expressed in percentage with continuous compounding. The sample period runs from January 4, 1999 to May 24, 2011, totaling to 3174 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. $\hat{\rho}(n)$ *is the n-day lag autocorrelation.*

5. Empirical Comparison of the Estimation Methods

As described in Section 3, we evaluate both the level (Section 5.1) and the conditional volatility (Section 5.2) of the extrapolated long term rates as proxied by the 30-year swap rate. The conditional volatility is evaluated both on the basis of the contemporaneous fit of the proxies with the observed rates as on the forecasting ability of the different procedures.

5.1. Fitting the Term Structure of Interest Rates

Although we put less emphasis on the in-sample fitting errors, we still report them (Subsection 5.1.1). Subsection 5.1.2 discusses the accuracy and economic attractiveness of the extrapolated long rates.

5.1.1. In-Sample Fitting Errors

Table 3 shows the in-sample MAE for all estimation procedures. Thanks to the additional parameters in the Svensson model and the flexibility of a 2-dimensional grid search, the 2dimGS returns the lowest in-sample MAE in 18 out of 22 cases. This results across all maturities in an average MAE of a mere 1.94 basis points. Taking the GS - that Nelson and Siegel (1987) originally used - as our benchmark, the 2dimGS MAE shows an improvement of respectively 10.96 basis points and 2.67 basis points for the 1-week and the 10-year rates. The overall average MAE decreases by 61%.

Table 3 In-sample MAE between estimated rates from Svensson-type models and market data (in basis points)

	GS	GS-RR	2×GS	2dimGS	2×GS -RR	2dimGS-RR	2×GS/GS-RR	2dimGS/GS-RR
1 Week	15.32+	14.40	4.94	4.36*	4.90	7.41	4.98	6.79
1 Month	10.36+	9.66	5.17	4.94*	5.58	6.05	5.53	6.43
2 Months	5.65+	5.32	3.60	3.38*	3.92	4.06	3.87	4.29
3 Months	3.86	4.15+	2.72*	2.75	3.03	2.94	3.00	3.17
4 Months	3.54	3.85+	2.36	2.09*	2.28	2.29	2.28	2.24
5 Months	3.94	4.22+	2.21	2.01*	2.05	2.36	2.07	2.24
6 Months	4.70	4.87+	1.90	1.84*	2.10	2.72	2.10	2.68
7 Months	4.63	4.70+	1.49*	1.56	1.72	2.54	1.68	2.49
8 Months	4.70+	4.68	1.12*	1.26	1.51	2.43	1.41	2.37
9 Months	4.96+	4.84	1.23	1.02*	1.72	2.44	1.57	2.50
10 Months	4.98+	4.78	1.69	1.24*	2.08	2.63	1.93	2.74
11 Months	5.11+	4.85	2.27	1.84*	2.65	2.96	2.47	3.18
12 Months	5.45+	5.14	3.03	2.63*	3.38	3.49	3.18	3.80
2 Years	6.39	7.72+	4.21	2.99*	6.33	6.10	6.28	5.53
3 Years	5.67	7.42+	2.01	1.74*	4.63	5.61	4.59	4.46
4 Years	4.63	6.46+	0.90*	1.08	3.63	5.30	3.56	3.44
5 Years	3.55	5.07+	1.35	1.12*	2.89	4.72	2.79	3.13
6 Years	2.62	3.15+	1.51	1.22*	1.91	2.99	1.88	2.25
7 Years	1.33	1.21	1.19	0.90*	1.33	1.32	1.39+	1.19
8 Years	1.48	2.16	0.89	0.72*	2.15	2.87+	1.99	1.70
9 Years	2.51	4.28	0.64	0.61*	3.91	5.60+	3.33	3.28
10 Years	3.97	6.31	1.57	1.30*	5.64	8.27+	4.71	4.97
Average	4.97	5.42+	2.18	1.94*	3.15	3.96	3.03	3.40

*Note: The sample period runs from January 4, 1999 to May 24, 2011. The dataset used to estimate the parameters is composed by 1-week, 1- to 12-month, and 2- to 10-year spot rates. * This approach yields the lowest MAE for this time to maturity. + This approach yields the highest MAE for this time to maturity. For a summary of the estimation procedures, see Table 1.*

Only the GS-RR performs worse than Nelson and Siegel's original GS *in-sample*, due to the bias added by the ridge regression.

5.1.2. Out-of-Sample Fit and Fitting Errors

Based on the in-sample MAE, the 2dimGS seems to be the most accurate model to describe the yield curve. However, as ridge regression introduces bias into OLS, in-sample fit is not a 'fair' measure of comparison, not to mention that our main concern is on extrapolation of the long end of the yield curve. We use the same procedure as in Annaert et al. (2013) to generate 30-year swap rates and compute the MAE of each estimation procedure.

Table 4 summarizes the descriptive statistics on extrapolated 30-year swap rates based on all eight models. For all the models, the extrapolated 30-year swap rates seem to be in line with the market observed ones.

Table 4 Descriptive statistics on extrapolated 30-year swap rates (in percentage)

Model	Mean	Std. Dev.	Skew.	Kurtosis	Min.	Max.
Observed rates	4.61	0.79	-0.03	2.23	2.53	6.18
GS	4.69	0.82	0.13	2.17	2.55	6.92
GS-RR	4.58	0.77	-0.01	2.15	2.50	6.11
2×GS	4.88	0.86	0.19	1.99	2.28	7.39
2dimGS	4.74	0.80	0.17	2.10	2.77	7.27
2×GS-RR	4.52	0.71	0.34	2.16	2.92	6.06
2dimGS-RR	4.44	0.80	0.01	2.58	2.14	6.12
2×GS/GS-RR	4.62	0.73	0.13	2.02	2.92	6.11
2dimGS/GS-RR	4.57	0.76	0.06	2.20	2.50	6.12

Note: Interest rates are expressed in percentage with continuous compounding. The sample runs from January 4, 1999 to May 24, 2011, totaling 3174 days. The 30-year swap rates based on various estimation procedures are summarized in this table. In the first row, the descriptive statistics on the observed market rates are reported.

Table 5 shows the MAE, the mean squared errors and the bias decomposition in the extrapolated 30-year swap rates. The original Nelson and Siegel grid search shows a MAE of 27.01 basis points for the 30-year swap rate. The inclusion of an extra hump factor in the Nelson-Siegel model generally pays off in terms of MAE. Caution is however warranted. The MAE for the long rate using the 2×GS procedure is almost double the size of the GS. Augmenting the GS with a ridge regression seems to alleviate that problem. Notice that the 2dimGS/GS-RR procedure has MAE of 11.61 basis points for the long rates, which implies an improvement of more than 50% vis-à-vis the GS.

Table 5 also presents the mean squared errors (MSE) and the bias resulting from the Theil-decomposition (1967). MSEs lead to the identical conclusion. The 2dimGS/GS-RR seems to be the best performing model in extrapolating the long end of the yield curve. The bias introduced by 2dimGS/GS-RR is less than both the 2dimGS and 2dimGS-RR.

Table 5 Out-of-sample fitting errors for 30-year swap rates

Model	GS	GS-RR	2×GS	2dimGS	2×GS-RR	2dimGS-RR	2×GS/ GS-RR	2dimGS/ GS-RR
MAEs (bps)	27.01	13.51	47.74+	18.12	23.29	20.13	14.50	11.61*
MSEs (10^{-4})	23.80	3.49	43.21+	8.31	8.25	6.72	4.13	2.56*
Bias (10^{-4})	7.93	2.24	27.19+	13.03	9.12	16.36	0.96*	4.02

*Note: Out-of-sample mean absolute errors, the mean squared errors and the bias decomposition based on mean squared errors between produced data from Nelson-Siegel and Svensson and empirical data are presented. The dataset used to estimate the parameters is composed by 1-week, 1- to 12-month, and 2- to 10- year spot rates. * This approach yields the lowest value for this time to maturity. + This approach yields the highest value for this time to maturity.*

5.2. Fitting and Forecasting the Term Structure of Volatilities

5.2.1. Fitting and Forecasting Errors

After analyzing the impact of these estimation procedures on the yield curve itself, we also examine their ability to fit and forecast the term structure of interest rate volatilities. This section discusses the behavior of the long rate volatility. We first calculate the mean absolute errors between the time series of conditional volatilities from market observed and model implied 30-year swap rates. Next, we forecast one-day ahead conditional volatilities.

Table 6 presents the MAE of the conditional volatility estimated on extrapolated 30-year rates from the Nelson-Siegel-Svensson-type models and on 1-day-ahead forecast of the conditional volatility of 30-year swap rates.

Focusing again on the mean absolute fitting errors and using the GS as benchmark, the MAE amounts to 20.14 basis points for the 30-year swap rate. 2×GS yields the both the highest fitting errors and forecasting errors. 2dimGS/GS-RR, GS-RR and 2×GS/GS-RR strongly reduce the MAE. Here the method 2dimGS/GS-RR has the lowest fitting errors with an MAE of only 9.41 basis points.

Table 6 MAE's on conditional volatility fit and forecasts (in basis points)

Model	GS	GS-RR	2×GS	2dimGS	2×GS-RR	2dimGS-RR	2×GS/GS-RR	2dimGS/GS-RR
Fitting errors	20.14	12.06	73.56+	25.18	43.91	24.45	13.59	9.41*
Forecasting errors	30.39	20.62	55.27+	28.15	34.41	28.00	25.35	13.85*

*Note: MAE's between the conditional volatility estimated on extrapolated 30-year rates derived from Svensson-type models and their empirical counterparts are reported as fitting errors. MAE's between one-day-ahead conditional volatility forecasts estimated on extrapolated 30-year rates derived from Svensson-type models and their empirical counterparts. The dataset used to estimate the parameters is composed by 1-week, 1- to 12-month, and 2- to 10-year spot rates. * This approach yields the lowest value for this time to maturity. + This approach yields the highest value for this time to maturity.*

Also in Table 6, we evaluate the ability of the estimation procedures to forecast one-day-ahead dynamics. We start by estimating a separate student-*t*-EGARCH(1,1) model for the 30-year swap rates on a training period of 2000 days. Next, we forecast the conditional volatilities one-day ahead and we compare the forecasts between the volatility of the extrapolated and observed rates on the same day. We have 3174 days in our sample, which means that the forecasts run from day 2001 to day 3174, totaling 1173 days. Every day we re-estimate the EGARCH models using expanding windows.

The GS shows a MAE of 30.39 basis points for the 30-year swap rate. GS-RR only improves the forecasting ability of the 30-year swap rate to 20.62 basis points. Here 2dimGS/GS-RR has the lowest MAE of 13.85.

5.2.2. Is the 2dimGS/GS-RR method statistically dominated by its competitors?

We started our quest with seven challengers for the grid search originally suggested by Nelson and Siegel (1987). Looking back to Table 3, among the methods withheld, the 2dimGS procedure produces the lowest MAE among all 8 methods with an average MAE over all maturities of only 1.94 basis points. This can be explained by the flexibility this method provides. From Table 5 we recall that the MAE of the 2dimGS/GS-RR procedure is 11.61 basis points for extrapolated 30-year swap rates. The second lowest MAE is generated by GS-RR, which is almost half of that of GS. Here our results are in line with Annaert et al. (2013) that ridge regression conditioned upon grid search can improve extrapolation quality of the Nelson-Siegel style models. Turning to the procedure's ability to match and forecast the volatility, once again 2dimGS/GS-RR yields the lowest fitting errors in both cases. 2dimGS method, the most flexibility method of all, has a mean absolute fitting errors that is almost 16 basis points higher than that of 2dimGS/GS-RR. This leads us to advance the 2dimGS/GS-RR procedure as the 'optimal' estimation method for Svensson-type term structure models when extrapolation of long-term interest rates is of interest.

Formally, we follow Ashley (1998) to run an MAE ratio test to measure the performance of these methods in capturing the long end of the yield curve. To test whether two series of MAE are statistically different from each other, we compute:

$$\alpha = \frac{MAE_{Method\,1}}{MAE_{Method\,2}} \tag{3}$$

where α is the MAE ratio. If α is significantly less than 1, then we consider Method 1 to be better than Method 2. Since the methods that we propose are not independent from each other and since we observe clustering in the MAE, we block-bootstrap absolute errors in pairs to calculate the Ashley test statistic. [4] To do so, we first divide the block-resampled time series from Method 1 and Method 2 by $MAE_{Method\,1}$ and $MAE_{Method\,2}$. Then we compute the MAE ratios. We resample 50 000 times and draw the distribution of the MAE ratio to compute the empirical *p*-value. In Table 7, negative spreads (i.e. $\alpha > 1$in the table) mean that the 2dimGS/GS-RR method has a lower MAE than the challenger. From Table 7 it is clear that 2dimGS/GS-RR significantly dominates other models in all tests. Thus this method seems to be able to extrapolate long-term rates as well as fit and forecast their conditional volatilities.

[4] The bootstrapped results are robust using various block sizes ranging from 15 up to 120. The optimal block size based on the algorithm of Politis and White (2004) varies from 80 to 120. Here we report results based on a block size of 50.

Table 7 Bootstrapped Ashley test (*p*-values in percentage, Spread in bps)

Challenger	GS	GS-RR	2×GS	2dimGS	2×GS-RR	2dimGS-RR	2×GS/GS-RR
Panel A: The spread between the extrapolation errors							
Spread	-15.40	-1.90	-36.13	-6.51	-11.68	-8.52	-2.89
α	2.33	1.16	4.11	1.56	2.01	1.73	1.25
p-value	(0.00)	(1.82)	(0.00)	(0.00)	(0.00)	(0.00)	(0.32)
Panel B: The spread between the student-*t*-EGARCH fitting errors							
Spread	-10.73	-2.65	-64.15	-15.78	-34.50	-15.04	-4.18
α	2.14	1.28	7.82	2.68	4.67	2.60	1.44
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Panel C: The spread between the student-*t*-EGARCH prediction errors							
Spread	-16.53	-6.77	-41.42	-14.30	-20.55	-14.15	-11.50
α	2.19	1.49	3.99	2.03	2.48	2.02	1.83
p-value	(0.01)	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Note: We perform a blocked bootstrap (block size = 50) with 50 000 runs. The spread between the 2dimGS/GS-RR and the challenger is reported in the first row. The second row reports the α as defined in Equation (3), and the empirical p-values (in percentage) are reported in brackets.

6. Conclusion

Annaert et al. (2013) propose to use ridge regression to estimate the Nelson-Siegel model in order to alleviate the multicollinearity problem embedded in the model. Their findings suggest that ridge-regression based algorithm has superior power in extrapolating long-term rates. This paper explores the use of ridge regression as an alternative for curve-fitting the extended Nelson-Siegel model, the Svensson model. We find that, the ridge regression conditioned on both Nelson-Siegel and Svensson (2dimGS/GS-RR) has consistently good performance in fitting the long end of the term structure of the yield curve itself and its volatilities, and in forecasting one-day ahead conditional volatilities. This promotes 2dimGS/GS-RR as a good candidate among other Nelson-Siegel-Svensson estimation procedures to build the long end of the term structure of interest rates and the term structure of volatilities.

7. References

Annaert, J., Claes, A.G.P., De Ceuster, M.J.K., and Zhang, H., Estimating the Spot Rate Curve Using the Nelson-Siegel Model: A Ridge Regression Approach. International Review of Economics & Finance (forthcoming, 2013, DOI: j.iref.2013.01.005).

Ashley, R., A new technique for postsample model selection and validation. *Journal of Economic Dynamics and Control*, 1998, 22(5), 647 – 665.

Bank of International Settlements, Zero-coupon yield curves – technical documentation. BIS Papers 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*, 1995, 5(2), 53-64.

Bolder, D. and Stréliski, D., Yield curve modeling at the Bank of Canada. Bank of Canada Technical Report No. 84, 1999.

Cairns, A.J.G., and Pritchard, D.J., Stability of descriptive models for the term structure of interest rates with application to German market data. *British Actuarial Journal*, 2001, 7(3), 467 – 507.

de Pooter, M., Examining the Nelson-Siegel class of term structure models. Tinbergen Institute Discussion Paper IT 2007-043/4, 2007.

Díaz, A., Jareño, F. and Navarro, E., Term structure of volatilities and yield curve estimation methodology. *Quantitative Finance*, 2011, 11(4), 573 – 586.

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

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

Ferenczi, I. and Werner, R., Calibration of the Svensson model to simulated yield curves (internet). http://www.mathfinance.de/workshop/2006/papers/werner/slides.pdf, 2006 (accessed on July 5, 2010).

Gilli M., Grosse. S., and Schumann, E., Calibrating the Nelson-Siegel-Svensson model. COMISEF Working Papers Series WPS-031, 2010.

Gürkaynak, R.S., Sack, B. and Wright, J.H., The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 2007, 54(8), 2291 – 2304.

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

Litterman, R. and Scheinkman, J., Common factors affecting bond returns. *Journal of Fixed Income*, 1991, 1(1), 54-61.

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

Politis, D.N. and White, H., Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 2004, 23(1), 53-70.

Ramponi, A., and Lucca, K., On a generalized Vasicek-Svensson model for the estimation of the term structure of interest rates. IV Workshop di Finanza Quantitativa Torino, ICER, Torino, 30 – 31 Gennaio, 2003.

Svensson, L.E.O., Estimating and interpreting forward interest rates: Sweden 1992-1994. International Monetary Fund Working Paper No. 94/114, 1994.

Theil, H., *Economics and Information Theory*, 1967 (Rand McNally and Company: Chicago)

Appendix 1: Ridge Regression Implementation

To overcome OLS parameter instability due to multicollinearity, we implement ridge regression. 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 the small variance estimators are less sensitive to measurement errors. We therefore use the ridge regression and compute our estimates as follows:

$$\widehat{\boldsymbol{\beta}}^{*} = \left[\mathbf{X}'\mathbf{X} + k\mathbf{I}\right]^{-1}\mathbf{X}'\mathbf{y}, \tag{4}$$

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.

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 unbiased estimates. As a result, we implement a combination of the grid search and the ridge regression using the following steps:

1. Estimate the parameters using one of the proposed methods (GS, 2dimGS, or 2×GS) ;
2. Calculate the condition number conditional on the 'optimal' estimates of the shape parameters ;
3. Re-estimate the coefficients by using ridge regression only when the condition number is above 10. The size of the ridge constant is chosen using an iterative search procedure that finds the lowest positive number k that makes the recomputed condition number fall below the threshold. Specifically, we start with k = 0 and we iteratively re-compute the condition number after increasing the ridge constant by 0.001. We stop iterating when the recomputed condition number is lower than 10. By adding a small bias, the correlation between the regressors decreases and so does the condition number.