European Journal of Operational Research 219 (2012) 442–451

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Linearized Nelson–Siegel and Svensson models for the estimation of spot interest rates

Geneviève Gauthier [a], Jean-Guy Simonato [b,*]

[a] Department of Management Sciences, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7
[b] Department of Finance, HEC Montréal, 3000 Côte-Sainte- Catherine, Montréal, Canada H3T 2A7

ARTICLE INFO

Article history:
Received 22 November 2010
Accepted 2 January 2012
Available online 10 January 2012

Keywords:
Term structure of interest rates
Spot rate curves
Coupon bonds
Prior information
Linearization

ABSTRACT

Linearized versions of the Nelson–Siegel (1987) and Svensson (1994) models for the cross-sectional estimation of spot yield curves from samples of coupon bonds are developed and analyzed. It is shown how these models can be made linear in the level, slope and curvature parameters and how prior information about these parameters can be incorporated in the estimation procedure. The performance of the linearized models are assessed in a Monte Carlo setting and with a sample of US government bonds. The results reveal that the linearized models compare favorably to the original models in terms of parameter estimates stability, computing effort and prevalence of local optima.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The term structure of spot interest rates describes, at a given date, the yields of zero coupon bonds according to their maturities. These are essential inputs used for many purposes such as pricing derivatives, valuing investment projects or computing risk measures. Most bonds that are traded are coupon bearing bonds. Hence, except for short maturities, spot interest rates are unobserved and must be estimated from samples of coupon bond prices. Several approaches are available to perform this task. This paper focuses on the Nelson and Siegel (1987) and Svensson (1994) approaches for cross-sectional yield estimation from sample of coupon bonds.[1]

As mentioned in Sundaram and Das (2010), these approaches have many advantages. First, spot rate curves estimates obtained with such methods generate smooth forward curves, unlike the typical jagged curves obtained with the spline approach proposed in McCulloch (1971, 1975). Second, at a given date, these methods can use the information available in all bonds, unlike the bootstrapping approach which uses only a subset of bonds and that is sensitive to which bonds are used. A third advantage is the financial interpretation that can be attached to the parameters of these models. The parameters can be interpreted as the yield on a long-term zero-coupon bond, the slope, curvature and hump position of the spot rate term structure. This feature is of interest since principal components analysis of bond yields typically finds these factors to be the main drivers of the interest rate term structure.

There are also several drawbacks associated with the use of Nelson–Siegel–Svensson models. A first problem is the instability regarding the parameter estimates obtained with cross sections of coupon bonds. An illustration of this problem is provided in Fig. 1 which shows time series of parameters obtained from Gürkaynak et al. (2007) who have estimated the Nelson–Siegel–Svensson model on daily cross-sections of US coupon bonds.[2] The top graph in Fig. 1 presents the estimated short term rate, which is the sum of two parameters from the Nelson–Siegel–Svensson model. These estimates are unstable with sudden large downward and upward spikes. Such large movements are incompatible with what we expect from such a quantity. The bottom graph from this figure shows the long term rate estimates. Again, these estimates show much instability. Using monthly cross-section of US coupon bonds, our study also documents the same problems.

As mentioned in Sundaram and Das (2010), another problem with the Nelson–Siegel–Svensson modeling approach is the multiple local optima possibility. Because of the nonlinearities in the

* Corresponding author. Tel.: +1 514 340 6807.
E-mail address: jean-guy.simonato@hec.ca (J.-G. Simonato).

[1] In the recent academic literature, dynamic versions of these models have been recently examined in terms of in-sample fit and forecasting performance with zero-coupon bond samples. See for example Diebold and Li (2006) or Christensen et al. (2010). Unlike these papers, we focus here on a very common use of the Nelson–Siegel–Svensson models which simply fits a cross-section of yields from coupon bonds at a particular point in time. As reported by the Bank for International Settlements (2005), several central banks use Nelson–Siegel–Svensson models in this way to estimate zero-coupon yield curves (Belgium, Finland, France, Germany, Italy, Norway, Spain and Switzerland). Such an approach is also used frequently in the academic literature. See for example, Gürkaynak et al. (2007) or Elton et al. (2001).

[2] Their zero-coupon yield curves and parameter estimates are publicly available at http://www.federalreserve.gov/pubs/feds/2006.

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2012.01.004

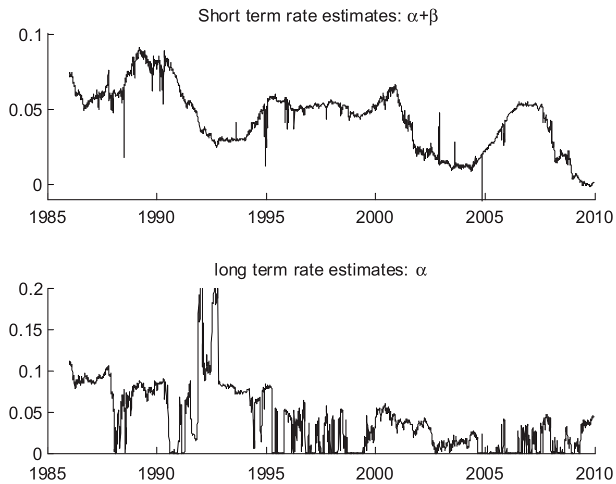

Fig. 1. Daily parameter estimates for the Svensson model from Gürkaynak et al. (2007).

parameters, it is not uncommon to find several optimum for a given cross-section of coupon bonds. Fig. 2 presents the results obtained from a cross section examined in this paper. Three different optimum are found, each one with a different shape for the term structure. The usual cure for the multiple optimum problem is to use several starting points. But with six parameters for the Svensson model, looking at n starting points involves n^6 non-linear estimations. Such a numerical search can be quite demanding.

We propose here modifications of the Nelson–Siegel–Svensson models that attempt to solve these two problems. Our modifications involve a partial linearization of the models coupled with the use of prior information for the parameters with a financial interpretation. The partial linearization reduces the number of parameters over which a non-linear search must be performed. This partial linearization also eases the ways with which prior information about the parameter values can be incorporated in the estimation procedure. The prior information takes the form of an expected value and a standard deviation. This information, even introduced with loose standard deviations, produces stable estimates that are in line with the financial interpretations of these parameters.

The paper is organized as follows. In Section 2, we present the original models and an assessment of the multiplicity of local optima problem with a US government coupon bond sample. Our linearized versions of these models are then described in Section 3. The approach for introducing prior information is examined in Section 4. Section 5 shows the results of simulation studies comparing the performance of the proposed algorithms. Finally, Section 6 examines estimates of the probability of not having found the global optimum. The stability of the parameter estimates obtained with the different methods is also examined. Section 7 concludes.

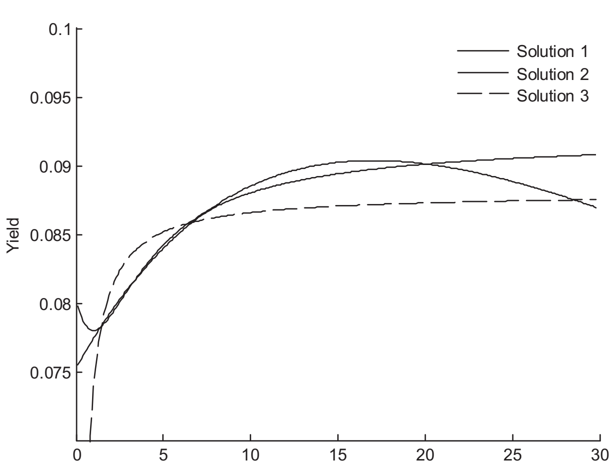

Fig. 2. Estimated zero-coupon yields from a sample of US government coupon bonds – Sept. 1990.

2. The Nelson–Siegel–Svensson models

We describe here the Nelson–Siegel (1987) and Svensson (1994) models and define the notation and concepts that will be used to introduce our linearized algorithms in the next sections. An assessment of the multiplicity of solution problem for these models is also provided here with monthly samples of US government bonds.

Nelson–Siegel (1987) specify the spot interest rates as a flexible functional form given by[3]:

$$y(T;\theta) = \alpha + \beta\phi_1(T;\tau) + \gamma\phi_2(T;\tau), \tag{1}$$

where T is the maturity expressed in years and $\theta = (\alpha,\beta,\gamma,\tau)^\top$ is a vector of unknown constant parameters. Moreover, $\phi_1(T;\tau) = \frac{\tau}{T}\left(1 - e^{-\frac{T}{\tau}}\right)$ and $\phi_2(T;\tau) = \frac{\tau}{T}\left(1 - e^{-\frac{T}{\tau}}\right) - e^{-\frac{T}{\tau}}$. The financial interpretation of the parameters in θ can be uncovered by looking at the behavior of $\phi_1(T;\tau)$ and $\phi_2(T;\tau)$ as T varies. For a fixed positive value of τ, $\phi_1(\bullet)$ tends to one as $T \to 0$ and then gradually decreases to zero as T increases. $\phi_2(\bullet)$ starts at zero, increases gradually and then decreases to zero again as T is increased furthermore. It is thus clear that, as the maturity increases to infinity, $y(T;\theta)$ tends to α. As such, α can then be interpreted as the long term yield. Finally, as the maturity goes to zero, the yield converges to $\alpha + \beta$. This combination of parameters is thus interpreted as the short-term rate. This allows β to be interpreted as the slope of the term structure which corresponds to the difference between the short-term and the long term rates. The parameter τ determines the position of the hump in the spot rate curve while γ controls the importance of this hump. In a cross section study, Diebold and Li (2006) found a large correlation between γ and what they called the curvature which they defined as the difference between the slope of the middle and long term rates and the slope of the short and middle rates.

Svensson (1994) proposed an extension of the Nelson–Siegel (1987) model by allowing for an additional hump position in the spot rate curve with a parameter controlling for the sensitivity to this hump. For this model, the yields are thus written as:

$$y(T;\theta) = \alpha + \beta\phi_1(T;\tau) + \gamma\phi_2(T;\tau) + \delta\phi_3(T;\upsilon)$$

with $\phi_3(T;\upsilon) = \frac{\upsilon}{T}(1 - e^{-\frac{T}{\upsilon}}) - e^{-\frac{T}{\upsilon}}$ and $\theta = (\alpha,\beta,\gamma,\delta,\tau,\upsilon)$. This additional component makes it easier to fit term structure shapes with more than one local maximum or minimum along the maturity dimension.

2.1. Parameter estimation: The zero-coupon bond case

When dealing with a sample of zero-coupon bond yields, because we do not expect a perfect fit, we write the observed yields as:

$$y_{obs}(T_i) = y(T_i;\theta) + \varepsilon_i,$$

for $i \in \{1,2,\ldots,n\}$ where $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)^\top$ is a random vector with expectation $\mathrm{E}[\varepsilon] = \mathbf{0}_{n\times 1}$ and covariance matrix $\mathbf{\Sigma}_\varepsilon = \mathrm{E}[\varepsilon\varepsilon^\top] = \sigma_\varepsilon^2\mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix and n is the number of different

[3] The formulation found in Diebold and Li (2006) is used here. This formulation is equivalent to the original Nelson–Siegel (1987) but is easier to interpret.

maturities observed in the sample. Introducing an error term is justified since yields to maturity of zero-coupon bonds are usually estimated from a sample of coupon paying bonds using bootstrapping techniques. This is usually done at the cost of some approximation errors.

For the case of the Nelson–Siegel model, the unknown parameter vector $\theta = (\alpha, \beta, \gamma, \tau)^\top$ is then chosen to minimize the sum of squared errors defined as:

$$Q(\alpha, \beta, \gamma, \tau) = \sum_{i=1}^{n} (\alpha + \beta\phi_1(T_i; \tau) + \gamma\phi_2(T_i; \tau) - y_{obs}(T_i))^2.$$

Because $Q(\alpha,\beta,\gamma,\tau)$ might have several local optima, the design of a robust optimization procedure is important. A naive algorithm allowing the exploration of the parameter space can be designed as follows:

1. Construct a four-dimensional hyperbox $[\underline{b}_\alpha, \bar{b}_\alpha] \times [\underline{b}_\beta, \bar{b}_\beta] \times [\underline{b}_\gamma, \bar{b}_\gamma] \times [\underline{b}_\tau, \bar{b}_\tau]$ defining the acceptable parameter space over which a search will be performed.
2. Generate W random starting points $\Theta = (A, B, \Gamma, \Upsilon)$ according to a uniform distribution: if u_1, u_2, u_3 and u_4 are independent uniform random variables on the interval $[0,1]$ then $A = \underline{b}_\alpha + (\bar{b}_\alpha - \underline{b}_\alpha)u_1, B = \underline{b}_\beta + (\bar{b}_\beta - \underline{b}_\beta)u_2, \Gamma = \underline{b}_\gamma + (\bar{b}_\gamma - \underline{b}_\gamma)u_3$ and $\Upsilon = \underline{b}_\tau + (\bar{b}_\tau - \underline{b}_\tau)u_4$.
3. Evaluate the objective function $Q(\alpha,\beta,\gamma,\tau)$ for each of the W random starting points.
4. Perform v numerical minimizations of Q initialized with the v best starting points for which the objective function Q is found to be the smallest in step 3.

We will refer to this procedure as the naive algorithm. In this setting, W should be relatively large to increase the chances of reaching the global optimum.

As suggested in Nelson–Siegel (1987) and Diebold and Li (2006), if τ is fixed at some specific value τ_0, Eq. (1) becomes a linear function of $\phi_1(T;\tau_0)$ and $\phi_2(T;\tau_0)$ and can be rewritten as:

$$\mathbf{y}_{obs} = \mathbf{X}_{\tau_0}\theta_{\tau_0} + \varepsilon,$$

where $\theta_{\tau_0} = (\alpha_{\tau_0}, \beta_{\tau_0}, \gamma_{\tau_0})^\top$ with

$$\mathbf{X}_{\tau_0} = \begin{pmatrix} 1 & \phi_1(T_1;\tau_0) & \phi_2(T_1;\tau_0) \\ \vdots & \vdots & \vdots \\ 1 & \phi_1(T_n;\tau_0) & \phi_2(T_n;\tau_0) \end{pmatrix} \quad \text{and} \quad \mathbf{y}_{obs} = \begin{pmatrix} y_{obs}(T_1) \\ \vdots \\ y_{obs}(T_n) \end{pmatrix}.$$

Therefore, the objective function $Q_{\tau_0}(\alpha_\tau, \beta_\tau, \gamma_\tau) = Q(\alpha, \beta, \gamma, \tau_0)$, viewed as a function of three parameters, may be minimized analytically using a least squares estimator

$$\hat{\theta}_{\tau_0} = \left(\mathbf{X}_{\tau_0}^\top \mathbf{\Sigma}_\varepsilon^{-1} \mathbf{X}_{\tau_0}\right)^{-1} \left(\mathbf{X}_{\tau_0}^\top \mathbf{\Sigma}_\varepsilon^{-1} \mathbf{y}_{obs}\right).$$

Consequently, the initial objective function $Q(\alpha,\beta,\gamma,\tau)$, with four arguments, may be reduced to a single argument function $q(\tau) = Q(\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\gamma}_\tau, \tau)$. Adapting the naive algorithm to the above modification reduces the problem to a one-dimensional random grid search with an optimization algorithm searching over τ only.

The above naive algorithm can be easily adapted to the Svensson (1994) model. The algorithm then becomes a random search and optimization over six arguments. Again, as for the Nelson–Siegel case, it is possible to reduce the dimension of the problem by re-expressing the initial objective function. This would reduce the original problem to a two-dimensional random grid search with an optimization algorithm searching over τ and υ.

2.2. Parameter estimation: The coupon bond case

Let $B(m,c,T)$ denote the value of a coupon bond with a face value of 100 dollars where m is the number of remaining coupons, c is the annual coupon rate and T is the maturity expressed in years. The Nelson–Siegel bond price can be defined as

$$B(m_i, c_i, T_i; \theta) = \sum_{j=1}^{m_i} c_i^{(j)} \times P(T_i^{(j)}; \theta), \tag{2}$$

where $P(T;\theta) = e^{-y(T;\theta)\times T}$ is the discount factor. The dates $T_i^{(1)} < \ldots < T_i^{(m_i)} = T_i$ are the time to maturity (in number of years) of the coupons and the cash flow of the ith bond at time $T_i^{(j)}$ (assuming that the coupons are paid twice a year) is $c_i^{(j)} = 100\left(\frac{c_i}{2}\mathbf{1}_{j<m_i} + (1+\frac{c_i}{2})\mathbf{1}_{j=m_i}\right)$. The indicator function $\mathbf{1}_A$ is worth 1 if condition A is satisfied and zero otherwise. Because we do not expect a perfect fit between the Nelson–Siegel prices and the observed ones, we introduce an error term:

$$B_{obs}(m_i, c_i, T_i) = B(m_i, c_i, T_i; \theta) + \varepsilon_i, \tag{3}$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ is a random vector with expectation $E[\varepsilon] = \mathbf{0}$ and a diagonal covariance matrix $\mathbf{\Sigma}_\varepsilon$.

The parameter vector, $\theta = (\alpha, \beta, \gamma, \tau)^\top$ for Nelson–Siegel and $(\alpha, \beta, \gamma, \delta, \tau, \upsilon)$ for Svensson, is usually estimated by minimizing the sum of squared errors

$$Q(\theta) = \sum_{i=1}^{n} \left[(B(m_i, c_i, T_i; \theta) - B_{obs}(m_i, c_i, T_i)) \times \frac{1}{D_i} \right]^2, \tag{4}$$

where D_i is the modified duration of the ith coupon bond. The naive algorithm may be used to obtain the parameter estimates.

As mentioned in Svensson (1994), Bolder and Stréliski (1999) and Gürkaynak et al. (2007), because of the nonlinear relationship between bond prices and yields, minimizing the unweighted squared pricing errors will not produce the best fit in yields. Weighting the prices by inverse durations approximately converts the pricing errors into yield fitting errors. As verified by Gürkaynak et al. (2007), fitting the above weighted price errors generates yield curve parameters that are virtually identical to parameters obtained from fitting yield errors. Alternatively, it would have been possible to directly fit the yields. However, this approach increases the computing costs by a large amount because of the additional numerical work required by the yield conversion.

2.3. The multiplicity of solution problem

Two problems can be obtained when carrying out the Nelson–Siegel or Svenson approaches. First, different parameter values may lead to similar yield curves. Such a result is possible because the model is estimated from coupon bonds with finite maturities. Over the observed maturities, different parameters can generate almost identical yield curve estimates. These yield curves would however be different if longer maturities were considered. This indetermination problem will be addressed in Section 4 where the use of prior information will be studied to favor solutions with financially plausible parameter values.

The second problem comes from the optimization procedure. The function to optimize is not guaranteed to be convex and may show several local optima. To illustrate this problem, we look at the distinct solutions found with a monthly sample of US government coupon bonds from the Lehman Brothers Fixed Income Database distributed by Warga (1998) (all bonds with optional features removed). The sample of monthly prices spans the January 1987 to December 1996 period. For each of the 120 months, 100 random starting points are used to estimate the Nelson–Siegel (1987) and Svensson (1994) models with a non-linear least-squares procedure with the weighted bond pricing errors defined in Eq. (4). The number of distinct solutions are recorded and summarized in Table S.1.[4] Because the ultimate goal is about estimating yield

[4] Tables, sections and figures labeled as S.x can be found in the Supplementary Material section.

curves, we use the estimated yields to discriminate between the various solutions. Two solutions are considered identical if the computed zero-coupon yields, for maturities of $0.1, 0.2, \ldots, T_{\max}$ years, all have absolute discrepancies smaller than a given threshold varying from 5 to 20 basis points. Note that $T_{\max}$ is the largest maturity observed for a given month.

For the Nelson–Siegel case, with a 10 basis points threshold, an average of 3.1 solutions was found with a maximum of 14 and a standard deviation of 2.1. For the Svensson case, the problem is more severe with an average of 7.4 solutions. Looking at the percentiles also shows that the problem is pervasive for both approaches and not due to a few months with many solutions. Table S.2 presents the detailed results obtained for the specific case of September 1990 for the Nelson–Siegel case. The solutions, in terms of parameter estimates and function values, are clearly different. Fig. 2 plots the estimated yield curves for the solutions reported in the table. As expected, these curves show different shapes and values for the short and long-term spot rates.

It is clear that such results raise the question about having observed or not the global optimum. A possible approach to answer this question is to explore more thoroughly the parameter space with other random starting points and optimizations. As it will be discussed in Section 6, an estimate of the probability of not having observed the global optimum can be computed using such a random search strategy. This estimate provides guidance for deciding if the search should be continued or stopped. But this approach, even with only four parameters, is costly in computer resources. For example, for the Nelson–Siegel case, the September 1990 sample requires on average 15 seconds of computer time to perform an optimization using function `lsqnonlin` from Matlab. Because several repeated non-linear optimizations may be required to get a reasonable estimate, such computer time is prohibitive given that several hundreds of estimations may be needed for a reasonable probability estimate. There is thus a clear need for lowering the computing time when estimating this model. For this purpose, next section proposes linearized versions of the Nelson–Siegel (1987) and Svensson (1994) models cutting down the number of parameters over which a non-linear search must be performed.

3. Linearized algorithms

3.1. One-step linearized algorithms

It is possible to adapt the dimension-reduction approach described earlier to obtain an approximation to the system described by Eq. (2). Our adaptation is based on a linearization of the discount factor $P(T;\theta) = \exp[-y(T;\theta)\times T]$ with the Taylor approximation $\exp(x) \cong 1 + x$. Such an approximation is accurate only if x is in the neighborhood of zero, which is not the case in the present context, especially for long maturities. Therefore, an additional parameter, φ, is introduced and the discount factor is rewritten as:

$$\exp[-\varphi T]\exp[(\varphi - y(T;\theta))T].$$

The goal is to obtain inside the second exponential an expression as close to zero as possible for all bonds in the sample so that the discount factor may be accurately linearized. Once the parameters φ and τ are fixed to some values φ_0 and τ_0, the linearization allows the analytical optimization of the other three parameters, leading to $\theta_{\varphi_0,\tau_0} = (\alpha_{\varphi_0,\tau_0},\ \beta_{\varphi_0,\tau_0}, \gamma_{\varphi_0,\tau_0})^\top$. Section S.1 shows, in the Nelson–Siegel framework, how this linearization introduced in the non-linear set of Eq. (3) leads to the following approximate system:

$$\mathbf{Y}_{\varphi_0} = \mathbf{X}_{\varphi_0,\tau_0}\theta_{\varphi_0,\tau_0} + \mathbf{e},$$

where $\mathbf{e} = (e_1,\ldots,e_n)^\top$ is a vector of zero-mean random noises with covariance matrix $\boldsymbol{\Sigma}_e = \mathrm{E}[\mathbf{e}\mathbf{e}^\top] = \sigma_e^2\mathbf{I}_n$, and the expressions for $\mathbf{Y}_{\varphi_0}$ and $\mathbf{X}_{\varphi_0,\tau_0}$ are available in Section S.1. This approximation is appropriate if $(y(T^{(j)};\theta) - \varphi)T^{(j)}$ is close to zero for any $T^{(j)}$. An intuitive candidate for φ can be, for example, the average yield-to-maturity. With this approximate system, the parameters can be estimated by minimizing the objective function $O(\alpha,\beta,\gamma,\tau,\varphi) = \sum_{i=1}^n e_i^2$. However, because this system is linear in $\theta_{\varphi_0,\tau_0}$, the five dimensional objective function $O(\alpha,\beta,\gamma,\tau,\varphi)$ may be reduced to a two-dimensional one. Given values for φ_0, τ_0, the least squares estimator for $\theta_{\varphi_0,\tau_0}$ is:

$$\hat{\theta}_{\varphi_0,\tau_0} = \left(\mathbf{X}_{\varphi_0,\tau_0}^\top\boldsymbol{\Sigma}_e^{-1}\mathbf{X}_{\varphi_0,\tau_0}\right)^{-1}\left(\mathbf{X}_{\varphi_0,\tau_0}^\top\boldsymbol{\Sigma}_e^{-1}\mathbf{Y}_{\varphi_0}\right), \tag{5}$$

which minimizes $O_{\tau_0,\varphi_0}(\alpha,\beta,\gamma)$. Therefore, the five dimensional objective function $O(\alpha,\beta,\gamma,\tau,\varphi)$ is reduced to:

$$o(\tau,\varphi) = O(\hat{\alpha}_{\varphi,\tau},\hat{\beta}_{\varphi,\tau},\hat{\gamma}_{\varphi,\tau},\tau,\varphi). \tag{6}$$

Adapting the naive algorithm to the above modification is then straightforward and reduces the problem to a two-dimensional random search with an optimization algorithm searching over τ and φ. This will be referred to as the one-step linearized algorithm.

The above procedure can be easily adapted to the Svensson (1994) model. For this case, the original optimization problem is reduced to a three-dimensional random search with an optimization algorithm searching over τ, υ and φ.

3.2. Two-step linearized algorithms

It is possible to further reduce the dimension of the problem. As noticed above, the additional parameter φ should be close to the yield to maturity for the approximation to work adequately. An obvious candidate for φ is therefore the average yield to maturity of the zero-coupon bonds. Building on this intuition, it is possible to further reduce the dimension of the problem and construct our algorithm to search mostly in the direction of hump location parameters. It will also consider only two different values for the extra parameter φ, explaining the reduction in dimension. The algorithm goes as follows for the Nelson–Siegel case:

1. Fix φ to an initial guess φ_0.
2. Apply the one-dimensional grid search to the system $o(\tau,\varphi_0) = O(\hat{\alpha}_{\varphi,\tau},\hat{\beta}_{\varphi,\tau},\hat{\gamma}_{\varphi,\tau},\tau,\varphi_0)$ to find an estimate $\tilde{\tau}$.
3. Compute $\tilde{\theta}_0 = (\hat{\alpha}_{\varphi_0,\tilde{\tau}},\hat{\beta}_{\varphi_0,\tilde{\tau}},\hat{\gamma}_{\varphi_0,\tilde{\tau}},\tilde{\tau})$ with $\left(\mathbf{X}_{\varphi_0,\tilde{\tau}}^\top\Sigma_e^{-1}\mathbf{X}_{\varphi_0,\tilde{\tau}}\right)^{-1}(\mathbf{X}_{\varphi_0,\tilde{\tau}}^\top\Sigma_e^{-1}\mathbf{Y}_{\varphi_0})$
4. Construct the yield to maturity curve $T \rightarrow y(T;\tilde{\theta}_0)$.
5. Compute the average yield $\bar{\varphi} = \frac{1}{M}\sum_{T=1}^{M} y(T;\tilde{\theta}_0)$ with M the largest maturity in the sample. Set $\varphi_0 = \bar{\varphi}$.
6. Perform steps 2–5.

This will be referred to as the two-step linearized algorithm. Again, this procedure can be easily adapted to the Svenson case with an optimization over τ and υ in each step.

As shown in numerical tests not reported in this study, the algorithm is not sensitive to the starting values for φ. This algorithm could also be generalized to perform iterations on the value of φ until convergence is reached for this parameter. However, we usually observe convergence in two or three iterations and the results are not improved by such a modification. Finally, modifying the algorithm to work with more than one value of φ could also be done. In this case, it would be required to split the maturity spectrum at arbitrarily specified time points to obtain the estimates. Again, our results did not show much sensitivity to such a modification.

The next section discusses how it is possible to take advantage of the financial interpretation assigned to the parameters (or linear combinations of the parameters) of the flexible functional form to introduce prior knowledge about these values.

4. Prior information

As mentioned in Section 2, it is often found that different parameter estimates lead to almost identical zero-coupon yield curves.[5] This raises the question about how financially plausible is each solution. Based on the information available to analysts on the date at which the estimation is performed, one of these solutions might be rejected *ex-post*, that is, after the estimation. Given that many different solutions may be obtained, this can quickly become a tedious procedure. A better approach would be to include this prior information directly in the estimation procedure. The occurrence of undesirable solutions, on an *ex-ante* basis, might then be reduced.

A first approach for incorporating the prior information in the estimation process would be to narrow the allowed parameter space during the non-linear optimization. However, it is possible that the expectations about the values are incorrect. A second approach, which is perhaps more interesting in our framework, is to introduce the prior information in a way reflecting this uncertainty. With linear systems, such an approach to include prior information is provided by the Theil–Goldberger estimator described in Goldberger (1964). This estimator uses information in the form of expected parameter values and the uncertainty about these expected values is characterized by standard deviations. The procedure for the Nelson–Siegel case, which can again be extended to the Svenson model, is detailed in Section S.2 of the supplementary material.

5. A Monte Carlo study

5.1. Assumptions

In this section, we perform a Monte Carlo study assessing the efficiency of the proposed approaches. Four different cases are considered: large ($N = 50$) and small ($N = 10$) sample sizes in combination with low and high noise variances. Each table will report the results obtained from 500 replications.

For the Nelson–Siegel case, the parameters are set to the average of the estimated parameters presented in Diebold and Li (2006). These average values of 0.075, −0.02, −0.0015, and 15 for α, β, γ, and τ arise from a monthly sample of zero-coupon yields between January 1987 to December 2000. For the Svensson case, δ, τ and υ are set to 0.05, 5 and 15.

For each replication, the coupon rates and the time to maturity are simulated using uniform random variables on the intervals [1%; 10%] and [1; 30] years respectively. Bond prices with annual coupons and face values of 100$ are first simulated without error terms. The yield to maturity of these bonds are then computed and i.i.d. zero mean normal error terms are added to these yields, which are then transformed back into prices. The standard deviations of these yield errors are set to 0.004 for the high noise case and to 0.00067 for the low noise case. These variance levels come from Elton et al. (2001) who report the average root mean squared errors (*ARMSE*) obtained by estimating the Nelson–Siegel model on samples of government and corporate coupon bonds. The low variance case corresponds to the *ARMSE* of government bonds while the high variance case is the *ARMSE* for BBB-rated corporate coupon bonds.[6]

The optimization steps are performed using the function `lsqnonlin` from Matlab with a tolerance of 1.0×10^{-6} for the differences in the function and parameter values between two successive iterations. The maximum number of function evaluations and maximum number of iterations are set to 10,000. During the numerical optimization, the interval for α, the long-term rate, is constrained to be in [0.01; 0.20]. The slope β which is the difference between the short and the long-term rates is assumed to be in [−0.5; 0.5]. The curvatures, γ and δ, are assigned to [−0.5; 0.5] while the position of the humps are assumed to be in [0.1; 30].

For the Bayesian approaches, the prior means are set to the true values of α, β, γ and δ. A prior about the short rate is also specified as the sum of α and β. For the prior standard deviations, loose and tight priors are considered. For the loose prior case, the standard deviations of these parameters are set to the standard deviation of the estimated parameters obtained in Diebold and Li (2006), which are around 0.015. The tight case sets the prior standard deviations to 0.0015 for these three parameters.

5.2. Numerical results

We present here summarized results for the parameters' estimation, computation time, and the yield curve's estimation. Three performance indicators measure the discrepancies between the Nelson–Siegel and the simulated bond prices or yields. The indicator labeled SSE1 corresponds to the objective function that has been optimized, evaluated at the optimal parameter values. This function differs for each algorithm. In the naive case, SSE1 is calculated from Eq. (4). In all other cases, it corresponds to Eq. (6) which is based on a linearized approximation of the true Nelson–Siegel bond price. To measure how these linearized methods perform as approximation to the original Nelson–Siegel model, SSE2 presents the square root value of Eq. (4) evaluated at the optimum. For the plain Nelson–Siegel case, SSE1 and SSE2 are the same. Because SSE2 is not the optimized function for the other approaches, it is expected to be larger than the plain Nelson–Siegel case. To measure the quality of the estimation procedures on the estimated yield curves, we report a root mean squared error measure based on the squared differences between the estimated zero-coupon yields and the true simulated yield curve for maturities of $1, 2, \ldots, 30$ years.

Table 1 examines the Nelson–Siegel cases with a sample size of 50 bonds and a low variance level for the random errors. For this case, all methods offer similar performances with respect to yield estimates. For the parameter estimation, the one-step, two-dimensional linearized algorithm with loose priors produces parameter estimates that are off from the true parameters when compared to those obtained with the other methods. The average convergence time of the Nelson–Siegel approach is roughly 14 to 20 times longer than the one of the linearized methods. Among these linearized methods, the slowest technique is the one-dimensional algorithm since it requires a two step one-dimensional non-linear optimization. The average estimates of φ are similar for all methods. As expected, with correctly specified priors, tight prior methods perform better.

To assess the robustness of the linearized algorithms, we perform several different Monte Carlo experiments. The detailed results of these experiments are available in the Supplementary material. We summarize here the main results of these experiments.

[5] An example is provided in the Supplementary material with Figure S.1 which plots the zero-coupon yield curves associated with three solutions found using random starting points on a sample of US government bonds in July 1995. These curves have similar shapes, but, as shown in Table S.3, the estimated parameters are fairly different.

[6] The ARMSE reported in Elton et al. (2001) are bond pricing errors. The yield error standard deviation corresponding to these bond pricing errors levels have been determined by simulation.

Table 1
Monte Carlo results for the Nelson–Siegel case: 50 bonds and low noise variance.

	α	β	γ	τ	φ	Time	SSE1	SSE2	rmse	dy1	dy5	dy10	dy30
Nelson–Siegel (500 valid cases)													
Mean	0.072	−0.017	−0.001	13.960		18.669	0.061	0.061	2.5	−1.3	0.6	−0.1	1.7
Std	0.007	0.007	0.015	8.586		6.942	0.009	0.009	1.2	5.1	1.9	1.8	5.1
Two-step, one-dimensional linearized Nelson–Siegel without priors (500 valid cases)													
Mean	0.070	−0.015	−0.003	11.365	0.062	0.977	0.061	0.061	2.5	−1.3	0.6	−0.2	1.1
Std	0.008	0.008	0.017	8.170	0.000	0.191	0.009	0.009	1.2	5.5	1.9	1.8	5.2
Two-step, one-dimensional linearized Nelson–Siegel with loose priors (500 valid cases)													
Mean	0.072	−0.016	−0.007	9.340	0.062	1.383	0.061	0.061	2.4	−1.1	0.6	−0.3	1.1
Std	0.003	0.003	0.006	5.934	0.000	0.226	0.009	0.009	1.1	4.7	1.9	1.8	4.9
Two-step, one-dimensional linearized Nelson–Siegel with tight priors (500 valid cases)													
Mean	0.074	−0.019	−0.002	13.842	0.062	1.352	0.062	0.062	1.7	0.9	0.0	−0.1	−1.6
Std	0.002	0.001	0.001	2.914	0.000	0.209	0.009	0.009	0.8	3.2	1.4	1.3	2.4
One-step, two-dimensional linearized Nelson–Siegel with loose priors (500 valid cases)													
Mean	0.060	−0.005	−0.004	4.149	0.066	0.983	0.060	0.385	2.9	0.1	0.1	−0.4	−1.0
Std	0.013	0.013	0.013	3.563	0.022	0.176	0.008	0.406	1.3	7.9	2.4	1.9	6.2
One-step, two-dimensional linearized Nelson–Siegel with tight priors (500 valid cases)													
Mean	0.075	−0.020	−0.002	17.268	0.067	0.933	0.062	0.102	2.0	0.7	−0.2	−0.1	−1.4
Std	0.002	0.001	0.000	5.019	0.008	0.289	0.009	0.071	1.0	3.9	1.6	1.5	4.3

True parameters: $\alpha = 0.075$, $\beta = -0.020$, $\gamma = -0.002$, $\tau = 15$. For the non-linear optimization, lower bounds and upper bounds for α, β, γ and τ have been set to 0.010, −0.500, −0.500, 0.100 and 0.200, 0.500, 0.500, 30.000. "SSE1" is the sum of squared errors obtained for the optimized function; "SSE2" is the sum of squared errors obtained for the Nelson and Siegel model with the estimated parameters; "rmse" is the root mean squared error of the estimated yields; "dy1", "dy5", "dy10" and "dy30" are the differences in basis points between the estimated and true yield for maturities of 1, 5, 10 and 30 years. 1000 random starting points have been evaluated with an optimization performed with the 10 best points and the lowest optimum value kept as the estimate. "valid cases" is the number of cases for which convergence was achieved. The Two-step, one-dimensional linearized algorithm has been initialized with φ equal to the prior mean for α plus half the prior mean of β.

As the variance level of the error terms is increased and/or as the sample size is decreased, clearer differences are emerging. Because the remaining three cases are qualitatively similar (large sample and high variance level, small sample with high and low variance levels), we only present the results for the extreme case of 10 bonds with large noise in Table S.4. As expected, the noise's variance influences the standard deviations and sum of squared yield errors when compared to their corresponding values in Table 1. For the spot rate estimates, the tight prior cases are performing the best and the plain Nelson–Siegel and two-step, one-dimensional linearized without priors are the worst. Similar results are obtained for the parameter estimates. The two-step, one-dimensional method without priors is highly biased and inefficient for β and γ. Loose priors are able to correct this situation. Indeed, the estimates show errors in the same range of those from the plain Nelson–Siegel case but with much smaller standard deviations. The two-step, one-dimensional algorithm, even if carried out with loose priors, is thus dominating the original Nelson–Siegel approach in terms of speed and precision.

Table S.5 shows the results of a Monte Carlo experiment for which the parameters α, β, γ and τ are generated randomly for each simulated data set, rather than being fixed like in the previous tables. Because the parameters are different for each of the 500 simulated data sets, the table reports the difference between the estimated and true parameters. The random parameters are generated according to a uniform distribution. The minimum and maximum values of the uniform distribution are taken from estimates of the one-dimensional linearized algorithm on our monthly sample of US government coupon bonds that will be presented next section. The intervals for the uniform simulations are $\alpha \in [0.0623, 0.1003]$, $\beta \in [-0.0524, -0.0004]$, $\gamma \in [-0.0841, 0.0307]$ and $\tau \in [0.1532, 30]$. 500 parameter sets are simulated. Again, the prior means for α, β and γ are set to the true values while the prior standard deviations are set to 0.015 for the loose case and 0.0015 for the tight case. The conclusions about the reliability of the algorithm are similar qualitatively to those from the previous tables. Our conclusions are thus not specific to parameter values chosen in the previous experiments.

Table S.6 shows the results of a Monte Carlo experiment for which the prior means are set to values different from the true parameter values. Here the prior mean for the short rate is set to 0.065 while the true short rate i.e. $\alpha + \beta$ is 0.055. The prior mean for α is set to 0.085 while the true value is 0.075 and the prior means for β and γ are set to 0 while the true values are −0.02 and −0.002. The results show that, in terms of yield estimates, the two-step, one dimensional algorithms with no prior and loose prior have precisions similar to those of the plain Nelson–Siegel algorithm. With tight priors, the precision deteriorates, as expected from the wrong priors. However, the linearized algorithms are about 20 times faster than the plain Nelson–Siegel case. The previous conclusions about the two-step algorithm with loose priors are thus robust to the use of incorrect prior means.

Table S.7 shows the results of a Monte Carlo experiment for the Svensson case with a large sample and low variance. This table examines here only the two-step algorithms. Unlike the Nelson–Siegel cases, the two-step algorithms are now two dimensional i.e. a non-linear search is performed on the two hump location parameters τ and υ. The linearized method with no prior does not perform adequately with large errors about yields and parameter estimates. The addition of loose priors obtains an estimator with a performance similar to the plain Svensson approach in terms of mean but with lower standard deviations for the yield in the short end of the curve. The average parameter values of the linearized algorithm with loose prior are also closer to the true values. The reduction in computing time is not as large as the one observed for the Nelson–Siegel case. There is still a substantial gain by a factor around 2 for the approach with priors.

6. The probability of observing the global optimum

In this section, we examine an estimate of the probability that a search strategy using random starting points has not observed all outcomes. This estimate focuses on the probability of still having unobserved outcomes after k independent trials. If there is a low probability of an unobserved outcome, there is a high probability that all maxima, including the global maximum, have been found.

In addition, the parameter and yield estimates obtained with each method are compared. These results will show the performance of the proposed methods in actual samples of coupon bond prices.

6.1. Theoretical framework and assumptions

Denote by U_k the probability of still having unobserved outcomes after k independent trials. Robbins (1968) has shown that the statistic $V_k = \frac{S_k}{k}$, where S_k is the number of singletons after k independent trials,[7] is a good estimate for the probability of still having unobserved outcomes after $k-1$ independent trials since $E(V_k - U_{k-1}) = 0$.

Using our monthly sample of US government coupon bonds from January 1987 to December 1996, 100 random starting points are used to estimate monthly parameter values and yield curves with the Nelson–Siegel and Svensson models with their corresponding linearized versions. In practical applications, it is easy for analysts to obtain prior values for the short rate from various sources. We therefore use, each month, a prior value for the short rate ($\alpha + \beta$) set according to the 3 month to maturity zero-coupon yield from the Federal Reserve Bank of St-Louis. Since the information about the short rate is often of good quality, we set the prior standard deviation for this linear combination of parameters to 0.0005, a reasonably tight value. With such a standard deviation, an interval of two standard deviations around the prior mean represents plus or minus ten basis points. The other prior used is a long run mean (α) set equal to the average bond yield in the sample for the month, with a loose prior standard deviation of 0.005. Finally, the slope and curvature are arbitrarily set to the average parameter values found in Diebold and Li (2006) with loose prior standard deviations of 0.01 for each. The optimization is performed with the same algorithms and settings described in the earlier Monte Carlo studies. The solutions are classified with a threshold value of 10 basis point for the maximum absolute difference between the estimated yields to maturity.

6.2. Numerical results: V statistics

Table 2 reports the results about the V statistics for the Nelson–Siegel case. On average, only 60% of the random starting points achieve convergence for the plain Nelson–Siegel case. For the linearized algorithms, 100% of the cases achieve convergence. The average of the V statistics computed each month is 0.03 for the Nelson–Siegel case with a standard deviation of 0.04 and a maximum of 0.29. The percentiles show that in at least 50 percent of the cases, the V statistic is 0.01 while it is smaller or equal to 0.04 in at least 75% of the samples. For most months, after 100 trials, there is thus a high probability that all local optima are found. For the other methods, most cases show a V statistic of zero. The mean number of distinct solutions is the highest for the Nelson–Siegel case. The one-dimensional algorithms show averages around two for the number of distinct solutions with, however, a small standard deviation. This suggests that the global optimum is usually obtained when two or three distinct solutions are found.

Other interesting findings relate to the size of the group of solutions corresponding to the lowest function value. Because uniform independent random numbers are used as starting values, this number allows to compute the probability of hitting the global optimum in a single trial, given that one hundred random starting points are used. For the Nelson–Siegel case, the average probability is 0.51 while it is over 0.90 for the two-step, one-dimensional linearized algorithms and 0.80 for the one-step, two-dimensional algorithm. The plain Nelson–Siegel also shows a greater standard deviation about this number than the other methods. The other statistics about these estimated probabilities also indicate clearly that the two-step, one-dimensional algorithms have a greater likelihood of hitting the global optimum in a single trial. Overall, the numbers from these tables suggest that the multiplicity of solution problem is much milder for the two-step one-dimensional linearized algorithms than for the other approaches.

Table S.8 from the supplementary material reports the results about the V statistics for the Svensson case. A first difference is the larger proportion of cases achieving convergence for the plain algorithm. The linearized algorithm, as in the Nelson–Siegel case, show 100% convergence. The V statistics are close to those obtained for the Nelson–Siegel case. Another notable difference is the lower estimates of the probabilities of hitting the global optimum in a single trial. However, the two-step algorithm with an average estimate of 0.76 dominates the plain Svensson with an estimate of 0.61.

6.3. Numerical results: Parameter estimates

As a by-product of these probability estimates, we have available the parameter and yield estimates associated with each method. The comparison between these estimates is interesting because it shows the similarities and differences in results obtained with each algorithm. These differences are best illustrated in Figs. 3 and 4 which show the time series estimates for the long term rate (α) and short term rate ($\alpha + \beta$) for the plain Nelson–Siegel and linearized Nelson–Siegel with priors. For the plain Nelson–Siegel approach, the estimated long term rates show many sudden downward spikes to values below 1%, an implausible behavior for this quantity during the sampling period under study. The linearized algorithm estimates with priors obtain stable long term rates estimates that are more in line with such a quantity. Similar observations apply to the short term rate estimates.

Table 3 reports summary statistics about the monthly estimates from each method. The average estimates for the short-term rate ($\alpha + \beta$) are similar for the Nelson–Siegel case and the methods with priors. These averages are close to the average risk-free yield of 0.0546 computed with the risk-free short-term yield series from the St-Louis Federal Reserve, which is used as priors. The standard deviation is however smaller for the linearized methods with priors, a result consistent with those of the Monte Carlo simulation. The standard deviations for these methods are close to the standard deviation computed from the risk-free yield time series used in the methods with priors. This standard deviation is 0.01641. From this number, we observe that the Nelson–Siegel method generates short-term rates showing too much variability when compared with the observed short-term rates.

The estimates of α for the Nelson–Siegel and the two-step algorithm with priors are roughly similar. The two-step without prior and one-step algorithm with priors stand out with a low estimate of the average infinite term rate and large standard deviations.

For the average estimates of β, the two-step, one-dimensional linearized approach with priors has again the lowest standard deviation. The average estimates and standard deviations of γ and τ are roughly equal for all methods. The more reliable method (two-step, one-dimensional with priors) is 40 times faster than the plain approach.

The average sum of squared errors of the optimized functions are different. The average values of SSE2 are close to the Nelson–Siegel function value for the two-step, one-dimensional methods, showing they produce reasonable approximation to the Nelson–Siegel model. The results about the computing times show that

[7] A singleton is a solution that has appeared only once in our k independent random trials.

Table 2
Summary statistics about the random search for optima for the Nelson–Siegel case with US government bonds – January 1987 to December 1996.

Method	Plain	2 Step 1 dim. No Bayes	2 Step 1 dim. Bayes	1 Step 2 dim. Bayes
Mean nb. of valid cases	59.76	100.00	100.00	100.00
Std. nb. of valid cases	33.45	0.00	0.00	0.00
Mean nb. of singleton	1.17	0.05	0.01	0.21
Std. nb. of singleton	1.70	0.22	0.09	0.41
Mean nb. of distinct solutions	3.10	1.63	1.51	2.73
Std. nb. of distinct solutions	2.07	0.54	0.52	0.80
Mean global optimum group size	50.86	93.96	95.15	84.38
Std. global optimum group size	31.27	8.24	7.81	10.18
Mean V statistic	0.03	0.00	0.00	0.00
Std. V statistic	0.04	0.00	0.00	0.00
Min. V statistic	0.00	0.00	0.00	0.00
Max. V statistic	0.29	0.01	0.01	0.01
25th Percentile	0.00	0.00	0.00	0.00
50th Percentile	0.01	0.00	0.00	0.00
75th Percentile	0.04	0.00	0.00	0.00
Mean of global optimum prob.	0.51	0.94	0.95	0.84
Std. of global optimum prob.	0.31	0.08	0.08	0.10
Min. of global optimum prob.	0.07	0.59	0.59	0.48
Max. of global optimum prob.	0.99	1.00	1.00	1.00
25th Percentile of global optimum prob.	0.17	0.91	0.93	0.79
50th Percentile of global optimum prob.	0.55	0.96	1.00	0.86
75th Percentile of global optimum prob.	0.80	1.00	1.00	0.91

This table reports statistics about the estimation results got with US government bond prices for the Nelson–Siegel model and the linearized algorithms. All models are estimated with a non-linear least squares algorithm initialized with 100 different random starting points each month from January 1987 to December 1996. Two solutions are considered identical if the computed zero-coupon yields, for maturities of $(0.1, 0.2, \ldots, T_{\max})$ years, all have absolute discrepancies smaller than a threshold value of 10 basis points. $T_{\max}$ is the largest maturity observed for a given month. The optimization are performed using the function `lsqnonlin` of Matlab with a tolerance of 1.0×10^{-6} for the differences in the function and parameter values between two successive iterations. The random starting parameter values are generated according to a uniform distribution on the following intervals: $\alpha \in [0.05\ 0.2]$, $\beta \in [-0.5\ 0.5]$, $\gamma \in [-0.5\ 0.5]$ and $\tau \in [0.01\ 30]$.

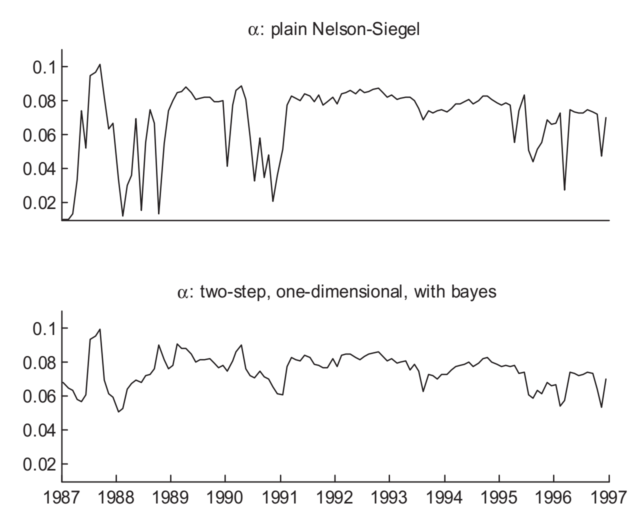

Fig. 3. Parameter estimates for the plain and linearized Nelson–Siegel models.

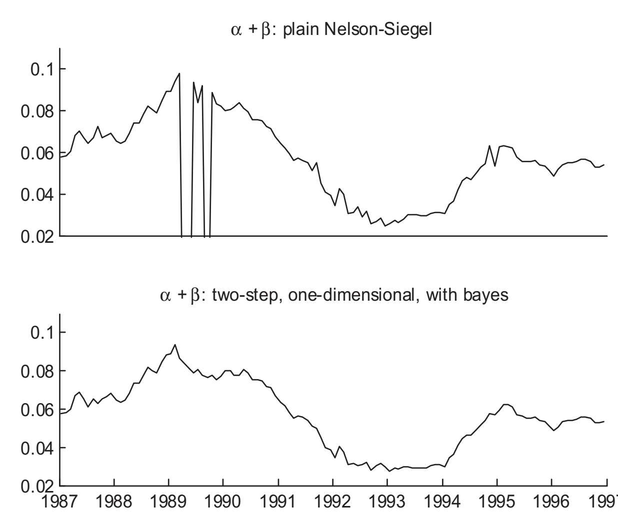

Fig. 4. Parameter estimates for the plain and linearized Nelson–Siegel models.

the Nelson–Siegel approach takes around 40 times the computing resources of the linearized algorithms.

Table 4 shows the yield estimates for different maturities. These results show that, for the maturities of the sample, the averages of the Nelson–Siegel and all linearized algorithms are close. For the shortest maturity, all estimators have an average close to the mean observed short-term spot rate. The linearized estimators with priors have, however, lower standard deviation for the shortest yield to maturity. For other maturities, the standard deviations of yield estimates are close.

This table also reports the average and standard deviations of the differences between the estimated yield and the estimated value of φ. As explained earlier, our linearized algorithms rely on a first-order Taylor series expansion of the exponential function around zero. The parameter φ is important in maintaining the argument $y(T;\theta) - \varphi$ close to zero. As the remainder of the first-order Taylor expansion is of order $(y(T;\theta) - \theta)^2 T^2$, we conclude from the means and the standard deviations exhibits in this table that the error is usually of magnitude 10^{-4}. To get a more precise assessment of the linearization error, we compute the difference between the linearized yields with a given value of φ and a given

Table 3
Summary statistics on parameter estimates from US government bonds for the Nelson–Siegel case – January 1987 to December 1996.

	α	β	γ	τ	φ	$\alpha+\beta$	SSE1	SSE2	Time
Nelson–Siegel									
Mean	0.0691	−0.0157	0.0325	7.2485	–	0.0533	0.58	0.58	17.16
Std	0.0204	0.0407	0.0569	7.2459	–	0.0305	0.41	0.41	17.95
Min	0.0100	−0.2175	−0.0507	0.1888	–	−0.1295	0.15	0.15	0.42
Max	0.1013	0.0651	0.2449	29.2960	–	0.0976	2.07	2.07	65.19
Two-step, one-dimensional linearized without priors									
Mean	0.0542	0.0002	0.0499	8.7580	0.0757	0.0544	0.57	0.66	0.47
Std	0.0523	0.0710	0.0965	9.9094	0.0096	0.0419	0.41	0.40	0.31
Min	−0.1526	−0.4333	−0.0643	0.1615	0.0578	−0.3514	0.14	0.15	0.11
Max	0.1013	0.2130	0.4535	30.0000	0.0958	0.0975	2.04	2.09	1.73
Two-step, one-dimensional linearized with priors									
Mean	0.0749	−0.0182	0.0207	6.4088	0.0757	0.0567	0.63	0.72	0.42
Std	0.0094	0.0204	0.0366	5.8194	0.0096	0.0178	0.47	0.46	0.25
Min	0.0507	−0.0571	−0.0579	0.1288	0.0578	0.0276	0.15	0.16	0.16
Max	0.0992	0.0142	0.0984	30.0000	0.0949	0.0938	2.24	2.41	1.84
One-step, two-dimensional linearized with priors									
Mean	0.0517	0.0047	0.0285	8.0864	0.0721	0.0564	0.59	75.62	0.63
Std	0.0347	0.0429	0.0524	8.3743	0.0322	0.0181	0.48	135.39	1.06
Min	−0.0568	−0.0566	−0.0621	0.1048	0.0100	0.0285	0.08	0.31	0.16
Max	0.0968	0.1087	0.1551	30.0000	0.1196	0.0974	2.29	680.72	10.70

This table reports statistics about the estimates got with US government bond prices for the Nelson–Siegel model and the linearized algorithms. All models are estimated with a non-linear least squares algorithm initialized with 100 different random starting points each months from January 1987 to December 1996. The optimization are performed using the function `lsqnonlin` of Matlab with a tolerance of 1.0×10^{-6} for the differences in the function and parameter values between two successive iterations. "SSE1" is the sum of squared errors obtained for the optimized function; "SSE2" is the sum of squared errors obtained for the Nelson and Siegel model with the estimated parameters. The random starting parameter values are generated according to a uniform distribution on the following intervals: $\alpha\in[0.05\ 0.2]$, $\beta\in[-0.5\ 0.5]$, $\gamma\in[-0.5\ 0.5]$ and $\tau\in[0.01\ 30]$.

Table 4
Summary statistics on yield estimates from US government bonds for the Nelson–Siegel case – January 1987 to December 1996.

	$T=0.1$	$T=1$	$T=5$	$T=10$	$T=20$	$T_{\max}$
Nelson–Siegel						
Mean $\hat{y}(T)$	0.0556	0.0614	0.0710	0.0761	0.0791	0.0793
Std $\hat{y}(T)$	0.0222	0.0167	0.0121	0.0103	0.0086	0.0073
Max $\hat{y}(T)$	0.0975	0.0965	0.0931	0.0964	0.0989	0.0997
Min $\hat{y}(T)$	−0.0453	0.0316	0.0479	0.0568	0.0609	0.0626
Two-step, one-dimensional linearized without prior						
Mean $\hat{y}(T)$	0.0565	0.0614	0.0710	0.0761	0.0791	0.0792
Std $\hat{y}(T)$	0.0263	0.0167	0.0121	0.0103	0.0086	0.0072
Max $\hat{y}(T)$	0.0974	0.0964	0.0932	0.0964	0.0989	0.0999
Min $\hat{y}(T)$	−0.1446	0.0318	0.0479	0.0566	0.0609	0.0627
Mean $[\hat{y}(T)-\hat{\varphi}]$	−0.0192	−0.0143	−0.0047	0.0004	0.0035	0.0035
Std $[\hat{y}(T)-\hat{\varphi}]$	0.0229	0.0103	0.0035	0.0010	0.0023	0.0039
Two-step, one-dimensional linearized with prior						
Mean $\hat{y}(T)$	0.0573	0.0612	0.0710	0.0761	0.0792	0.0797
Std $\hat{y}(T)$	0.0178	0.0166	0.0121	0.0103	0.0086	0.0074
Max $\hat{y}(T)$	0.0938	0.0952	0.0929	0.0959	0.0977	0.0983
Min $\hat{y}(T)$	0.0284	0.0318	0.0479	0.0566	0.0609	0.0627
Mean $[\hat{y}(T)-\hat{\varphi}]$	−0.0184	−0.0146	−0.0047	0.0004	0.0035	0.0039
Std $[\hat{y}(T)-\hat{\varphi}]$	0.0120	0.0101	0.0035	0.0010	0.0023	0.0037
One-step, two-dimensional linearized with prior						
Mean $\hat{y}(T)$	0.0570	0.0612	0.0709	0.0759	0.0794	0.0794
Std $\hat{y}(T)$	0.0180	0.0166	0.0120	0.0103	0.0086	0.0079
Max $\hat{y}(T)$	0.0973	0.0963	0.0928	0.0956	0.0993	0.1030
Min $\hat{y}(T)$	0.0286	0.0319	0.0479	0.0563	0.0614	0.0601
Mean $[\hat{y}(T)-\hat{\varphi}]$	−0.0187	−0.0146	−0.0048	0.0002	0.0036	0.0037
Std $[\hat{y}(T)-\hat{\varphi}]$	0.0121	0.0101	0.0034	0.0010	0.0023	0.0043

This table reports statistics about yields estimates got with US government bond prices for the Nelson–Siegel model and the linearized algorithms. All models are estimated with a non-linear least squares algorithm initialized with 100 different random starting points each months from January 1987 to December 1996. Maturities are in years and $T_{\max}$ is the largest maturity observed for a given month. The optimization are performed using the function `lsqnonlin` of Matlab with a tolerance of 1.0×10^{-6} for the differences in the function and parameter values between two successive iterations. The random starting parameter values are generated according to a uniform distribution on the following intervals: $\alpha\in[0.05\ 0.2]$, $\beta\in[-0.5\ 0.5]$, $\gamma\in[-0.5\ 0.5]$ and $\tau\in[0.01\ 30]$.

yield value. For example, using the average parameter estimate of 0.0757 for φ reported in Table 3 for the two-step linearized algorithm with priors, and the average yield value in Table 4 of 0.0556 for the 0.1 year maturity case, the approximate linearized yield can be computed as

$$y_{app}(T;\theta) = -\frac{\ln(1+\varphi T - y(T;\theta)T) - \varphi T}{T}$$

which obtains $y_{app}(0.1;\theta) = 0.05562$. This quantity is close to the target of 0.0556. The absolute values of the differences for other maturities are 1.01×10^{-4}, 5.44×10^{-5}, 8.02×10^{-7}, 1.21×10^{-4} and 2.10×10^{-4}. These differences are small and provide accurate approximations.

Finally, Tables S.9 and S.10 report the results for the Svensson case. The three methods obtain different results for the average estimated parameter values. Unlike the Nelson–Siegel case, the linearized method with no prior seems unreliable here (as it was the case with the Monte Carlo study) with counterintuitive average parameters values of −0.22 and 0.23 for α and β. Again, the linearized method with prior produces reasonable values for the short rate estimates, which are close in terms of mean and standard deviations to those from the short yield series used as priors. This is not the case for the plain Svensson method. The linearized method shows the smallest standard deviations for the parameter estimates and the lowest computing time, improving the plain Svensson case by an average factor around 25. As for the yield estimates, the three methods provides very close estimates, except for the very short end of the curve, which is estimated with the greatest precision with the linearized approach with prior.

7. Conclusion

Linearized Nelson–Siegel and Svensson algorithms for estimating spot interest rate term structures are developed and analyzed here. These algorithms retain the desirable features of the original approaches, such as the financial interpretation of the parameters, but are much faster to converge. An advantage of these linearized algorithms is the possibility to introduce prior information about some of the parameters of the model for which a financial interpretation is available. As shown by Monte Carlo experiments and an empirical study on samples of US government bonds, introducing such prior information enhance the precision with which the estimated spot rate curves can be obtained. Parameter estimates are also more stable and in line with their financial interpretations.

The probability of still having unobserved outcomes using a random starting point strategy is assessed for the Nelson–Siegel and Svenson models and the proposed linearized versions. Using monthly samples of US coupon bond prices from 1987 to 1996, the results show that multiplicity of solutions is a milder problem for the linearized algorithms. Our results suggest that, among the algorithms studied here, the two-step, one-dimensional linearized algorithm with prior information performs the best, even when used with very loose priors.

Further research could examine how these linearized models perform in terms of forecasting performance in the lines of Diebold and Li (2006). It could also be examined how they could be used in a filtering framework allowing the simultaneous estimation of the parameters and their dynamics.

Acknowledgements

Gauthier acknowledge the financial support of the National Science and Engineering Research Council of Canada (NSERC) and HEC Montréal. Simonato acknowledges the financial support of HEC Montréal. We thank Mathieu Boudreault and Diego Amaya for helpful comments and assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2012.01.004.

References

Bank for International Settlements, 2005. Zero-coupon yield curves: Technical documentation. Bank for International Settlements, Basle, Switzerland.
Bolder, D., Stréliski, D., 1999. Yield curve modelling at the Bank of Canada. Technical report no. 84, Bank of Canada.
Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D., 2010. The affine arbitrage-free class of Nelson–Siegel term structure models. Journal of Econometrics 164, 4–20.
Diebold, F., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364.
Elton, E.J., Gruber, M.J., Agrawal, D., Mann, C., 2001. Explaining the rate spread on corporate bonds. Journal of Finance 56, 247–277.
Goldberger, A., 1964. Econometric Theory. Wiley, New York.
Gürkaynak, R., Sack, B., Wright, J., 2007. The US treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 2291–2304.
McCulloch, J.H., 1971. Measuring the term structure of interest rates. Journal of Business 44, 19–31.
McCulloch, J., 1975. The tax adjusted yield curve. Journal of Finance 30, 811–830.
Nelson, R., Siegel, F., 1987. Parsimonious modeling of yield curves. Journal of Business 60, 473–489.
Robbins, H., 1968. Estimating the total probability of the unobserved outcomes of an experiment. The Annals of Mathematical Statistics 39, 256–257.
Sundaram, R., Das, S., 2010. Derivatives: Principles and Practice. McGraw-Hill, New York.
Svensson, L., 1994. Estimating and interpreting forward interest rates: Sweden 1992–1994. Centre for Economic Policy Research, Discussion Paper 1051.
Warga, A., 1998. Fixed Income Database. University of Houston, Houston, Texas.