In Search of the Best Zero Coupon Yield Curve for Nairobi Securities Exchange: Interpolation Methods vs. Parametric Models

Lucy Muthoni

Institute of Mathematical Sciences (IMS), Strathmore University, Nairobi, Kenya
Email: Lucymuthoni2003@gmail.com; lmuthoni@strathmore.edu

Abstract

We seek to determine which yield curve construction method produces the best zero coupon yield curve (ZCYC) for Nairobi Securities Exchange (NSE). The ZCYC should be differentiable at all points and at the same time, should produce a continuous and positive forward curve at all knot points. A decreasing discount curve is also expected from the resulting ZCYC, as an indication of monotonicity. For the interpolation method, we will use an improvement of monotone preserving interpolation method on $r(t)t$, while the Nelson and Siegel [1] model is the parametric model of choice.

This is because compared to other interpolation methods, the improvement of monotone preserving interpolation method on $r(t)t$ produces curves with the desirable trait of differentiability, while the Nelson-Siegel [1] model is shown to produce the best-fit results for Kenyan bond data. We compare the models' performance in terms of accuracy in pricing back the fixed-income securities. For this study, we use bond data from Central Bank of Kenya (CBK). The better of the two methods will be used for the Kenyan securities market and, consequently, the East African Securities markets.

Keywords

Interpolation Methods, Zero Coupon Yield Curves, Parametric Models, Nairobi Securities Exchange

1. Introduction

The objective of this paper is to establish the best yield curve model to be used in pricing of financial products at

the Nairobi Securities Exchange. We compare two models so far established to be the best in their respective categories, as shown by Muthoni Onyango and Ongati [2] and Muthoni Onyango and Ongati [3]. Among the Parametric Models used in construction of yield curves, Nelson and Siegel [1] model was found to be superior by Muthoni, Onyango and Ongati [3], while the improvement of monotone preserving interpolation on $r(t)t$ was found to be superior of the splines method studied by Muthoni, Onyango and Ongati [2]. This paper compares these two methods in terms of:

1) The accuracy of the yield curve using the set test statistics;
2) The smoothness of the curve;
3) The continuity and differentiability of the spot and forward curves;
4) Pricing back the fixed income securities;
5) The monotonicity of the yield curve, measured by the behavior of the discount curve.

2. Literature Survey

Many estimation methods for yield curves have appeared in literature over the years. Generally speaking, there are two distinct approaches to estimate the term structure of interest rates: the equilibrium model and the statistical techniques.

The first approach is formalized by defining state variables characterizing the state of the economy (relevant to the determination of the term structure) which are driven random processes and are related in some way to the prices of the bonds. It then uses no-arbitrage arguments to infer the dynamics of the term structure. Examples of this approach include: Vasicek [4], Dothan [5], Brennan and Schwartz [6] Cox, Ingersoll and Ross [7], and Duffie and Kan [8].

Unfortunately, in terms of the expedient assumptions about the nature of the random process driving the interest rates, the yield derived by those models have a specific functional form dependent only on a few parameters, and usually the observed yield curves exhibit more varied shapes than those justified by the equilibrium models.

In contrast to the equilibrium models, statistical techniques focusing on obtaining a continuing yield curve from cross-sectional coupon bond data based on curve fitting techniques are able to describe a richer variety of yield patterns in reality. The resulting term structure estimated from the statistical techniques can be directly put into the interest rate models such as the Hull [9], and Heath [10] models, for pricing interest rate contingent claims. Since a coupon bond can be considered as a portfolio of discount bonds with maturities dates consistent with the coupon dates, the discount bond prices can thus be extracted from the actual coupon bond prices by statistical techniques.[1] These techniques can be broadly divided into two categories: the splines (interpolation methods) and the parsimonious function forms (see Alper [11]).

2.1. Splines/Interpolation Models

Interpolation is a method of constructing new data points within the range of a discrete set of known data points (called knot points). The simplest method for interpolating between two points is by connecting them through a straight line. Some variations of linear interpolation are capable of ensuring a strictly decreasing curve of discount factors. However, all the variations of linear interpolation imply discontinuities in the forward rate curve.

In order to produce continuous forward rates curves, researchers introduced cubic methods of interpolation. An example of cubic interpolation algorithm is the cubic Hermite spline. Under cubic Hermite splines, the derivative of the data of each knot point is assumed to be known, and the interpolation function is required to be differentiable. Often, these derivatives will not be known, and will have to be estimated. One method for estimating these derivatives, described by de Boor [12] [13] as the Bessel method involves estimating the derivative though the use of a three point difference formula.

Unfortunately, all the traditional cubic methods are incapable of ensuring strictly positive forward rates, which are synonymous with non-decreasing discount factors, as shown by Hagan and West [14]. Furthermore, some cubic methods have an inherent lack of locality in the sense that a local perturbation of curve input data will cause ringing and cause changes in the data far away from the perturbed data point as shown by Anderson [15].

---

[1]Once the discount function, $P(t)$, is defined, the spot interest rate (the pure discount bond yield) can be computed by: $R(t) = \dfrac{-\ln P(t)}{t}$.

All variations of linear interpolations were seen to produce discontinuities in the forward rate curve, whilst all variations of cubic interpolations were seen to be incapable of ensuring strictly decreasing discount factors. Non-decreasing discount factors imply arbitrage opportunities, whilst discontinuous forward rates unacceptable from an economic perspective (unless the discontinuities occur on or around meetings of monetary authorities).

To counter this, a monotone convex interpolation method was developed, which it is claimed to be capable of ensuring a positive and (mostly) continuous forward rate curve Hagan and West [14]. This method proposed by Hagan and West [14], was specifically designed to interpolate yield curve data, and involves fitting a set of quadratic polynomials to a discrete set of estimated instantaneous forward rates. The method is designed such that $f(t)$ preserves the shape of the set of discrete forward rates. The monotone convex method was also seen to be capable of ensuring a strictly decreasing curve of discount factors. Unfortunately, the model depends heavily on an appropriate interpolation algorithm. In addition, it was discovered by du Preez [16] that there were specific conditions under which the interpolation function of the monotone convex interpolation would produce discontinuity of $f(t)$.

This led to the monotone preserving $r(t)t$ method of interpolation, introduced by Preez [16]. Essentially, this method involves applying cubic Hermite interpolation to the $r(t)t$ at the knot points i.e the values of $f(t)$ at the knot points, are estimated in manner which ensures positivity in $f(t)$. Constructing an interpolating algorithm capable of preserving the monotonicity of the discount factors, was thus sufficient for ensuring positive forward rates.

Monotone preserving $r(t)t$ method is capable of ensuring a positive and continuous forward rate curve, and was designed to preserve the geometry of $r(t)t$. Monotonicity in the discount factors implies monotonicity in the $r(t)t$ which is achieved by applying the work done in the field of shape preserving cubic interpolation, by authors such as de Boor [17], Carlson and Fritsch [18], Hyman [19] and Akima [20]. Apart from being an improvement of monotone convex method where it ensured positive forward rates, the monotone preserving $r(t)t$ method was also capable of ensuring continuity of $f(t)$.

In the study by du Preez [16], they found that the monotone preserving $r(t)t$ method to perfrom slightly better in terms of stability, and continuity of $f(t)$, compared to monotone convex method. This suggests that when bootstrapping, the monotone preserving $r(t)$ method could be the ideal method of interpolation. Unfortunately, monotone preserving method had the undesirable characteristic of not being differentiable at the knot-points.

Muthoni, Onyango and Ongati [2] introduced a new method of interpolation, which will be an improvement of monotone preserving $r(t)t$ interpolation method suggested by du Preez [16]. They was done by removing the non-differentiability at the knot points in the aforementioned method, which is created by use of Hyman monotonicity constraint, which enforces monotonicity.

2.2. Parsimonious Models

Parsimonious models specify parsimonious parameterizations of the discount function, spot rate or the implied forward rate. Moving from the cubic splines, Chambers [21] introduced the parsimonious function forms by considering an exponential polynomial to model the discount function. Nelson & Siegel [1] followed shortly thereafter by choosing an exponential function with four unknown parameters to model the forward rate of U.S Treasury bills. By considering the three components that make up this function, Nelson & Siegel [1] illustrated that it can be used to generate a variety of shapes for the forward rate curves and analytically solve for the spot rate. Moreover, the advantage of the classical Nelson-Siegel [1] model is that the three parameters may be interpreted as latent level, slope and curvatures factors. Diebold [22], Modena [23] and Tam & Yu [24] employed the Nelson-Siegel [1] interpolant to examine bond pricing with a dynamic latent factor approach and concluded that it was satisfactory.

Svensson [25] increased the flexibility of the original Nelson and Siegel [1] model by adding two extra parameters (Svensson [25] model) which allowed for a second "hump" in the forward rate curve. Later, Bliss [26] introduced the Extended Nelson-Siegel [1] method, which introduced a new appropriating function with five parameters by extending the model developed by Nelson & Siegel [1]. Bliss suggested that a six-parameter model can produce better results for fitting the term structure with longer maturities.

The Nelson-Siegel [1] model class has linear and non-linear parameters depending on the values assumed fixed. Due to this, these models have multiple local minima making model estimation difficult. Previous studies

have widely discussed the estimation of Nelson-Siegel [1] model class and they are: Bolder [27], Maria *et al.* [28], Gilli [29], Rezende [30], Rosadi [31], among others.

Muthoni *et al.* [3] estimated the Kenyan government bonds (KGBs) term structure of interest rates based on the parsimonious functions specifications , *i.e.* the four parameters Nelson-Siegel [1] model, the five parameters Svensson [25] model and the six parameters Rezende [30] model, known as Nelson-Siegel-Svensson model. The reason they chose the Nelson-Siegel family is that these models have substantial flexibility required to match the changing shape of the yield curve, yet they only use a few parameters. As noted by Diebold [22], they can be used to predict the future level, slope, and curvature factors for bond portfolio investments purposes. After comparing the Nelson-Siegel classes of models, Muthoni, Onyango and Ongati [3] found Nelson Siegel [1] to be the superior model.

3. Empirical Methodology

3.1. Model Selection

3.1.1. The Nelson-Siegel [1] Model

The Nelson-Siegel [1] model sets the instantaneous forward rate at maturity m given by the solution to a second order differential equation with unequal roots as follows:

$$f(m) = \beta_0 + \beta_1 \exp\left(\frac{-m}{\tau_1}\right) + \beta_2 \frac{m}{\tau_1} \exp\left(\frac{-m}{\tau_2}\right) \tag{1}$$

where $m > 0$. The model consists of four parameters: $\beta_0, \beta_1, \beta_2$ and τ_s. m is the time to maturity of a given bond. Equation (1) consists of three parts: A constant, an exponential decay functional and Laguerre function. β_0 is independent of m and as much, β_0 is often interpreted as the level of long term interest rates. The exponential decay function approaches zero as m tends to infinity and β_1 as m tends to zero. The effect of β_1 is thus only felt at the short end of the curve. The Laguerre function on the other hand approaches zero as m tends to infinity, and as m tends to zero. The effect of β_2 is thus only felt in the middle section of the curve, which implies that β_2 adds a hump to the yield curve.

The spot rate functions under the model of Nelson and Siegel [1] is as follows:

$$r(m) = \beta_0 + \beta_1 \left(\frac{\tau_1}{m}\right)\left[1 - \exp\left(\frac{-m}{\tau_1}\right)\right] + \beta_2 \left(\frac{\tau_2}{m}\right)\left[1 - \exp\left(-\frac{m}{\tau_2}\right)\right] - \beta_2 \exp\left(-\frac{m}{\tau_2}\right)$$

$$r(m) = \beta_0 + \beta_1 \left(\frac{\tau_1}{m}\right)\left[1 - \exp\left(\frac{-m}{\tau_1}\right)\right] + \beta_2 \left(\frac{\tau_2}{m}\right)\left[1 - \exp\left(-\frac{m}{\tau_2}\right)\left(\frac{m}{\tau_2} + 1\right)\right] \tag{2}$$

From Equation (2) it follows that both the spot and forward rate function reduce to $\beta_0 + \beta_1$ as $m \to 0$. Furthermore, we have $\lim_{m \to \infty} r(m) = \lim_{m \to \infty} f(m) = \beta_0$. Thus, in the absence of arbitrage, we must have that $\beta_0 > 0$ and $\beta_0 + \beta_1 > 0$.

Suppose we observe n zero coupon bonds, expiring at times $m_1, m_2, \cdots, m_n$. Let $p_1, p_2, \cdots, p_n$ denote the prices of these bonds. Note that p_i will imply the spot rate of interest corresponding to time m_i for $i = 1, 2, \cdots, n$. Let $r_1, r_2, \cdots, r_n$ denote these spot rates. If we assume that the values of τ_s are known, then the Nelson and Siegel [1] model reduces to a linear model, which can be solved using linear regression.

Define:

$$X = \begin{pmatrix} 1 & \frac{\tau_1\left(1 - e^{-m_1/\tau_1}\right)}{m_1} & \frac{\tau_1\left(1 - e^{-m_1/\tau_1}\right)}{m_1} - e^{-m_1/\tau_1} \\ 1 & \frac{\tau_2\left(1 - e^{-m_2/\tau_2}\right)}{m_2} & \frac{\tau_2\left(1 - e^{-m_2/\tau_2}\right)}{m_2} - e^{-m_2/\tau_2} \\ & \vdots & \vdots \\ 1 & \frac{\tau_n\left(1 - e^{-m_n/\tau_n}\right)}{m_n} & \frac{\tau_n\left(1 - e^{-m_n/\tau_n}\right)}{m_n} - e^{-t_n/\tau_n} \end{pmatrix} \quad Y = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} \quad B = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \tag{3}$$

We would like to obtain a vector B satisfying $XB = Y + \epsilon$. By using ordinary least squares (OLS) estimation, we can solve B as follows:

$$B = (X'X)^{-1} X'Y \tag{4}$$

Nelson and Siegel [1] suggested the following procedure for calibrating their model:
1) Identify a set of possible values for τ_s
2) For each of these τ_s estimate B
3) For each of these τ_s and their corresponding B's, estimate

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(r_i - \hat{r}_i)}{\sum_{i=1}^{n}(r_i - \overline{r}_i)} \tag{5}$$

4) The optimal λ and β are those associated with the highest value of R^2.

The above method of calibration is known as the grid-search method. Alternatively, non-linear optimization techniques can be used to solve all four parameters simultaneously. However, such networks are very sensitive to starting values, implying a high probability of obtaining local optima. The estimated parameters obtained by using the grid search method behave erratically over time, and have large variances. The problems resulting from multi-colinearity depend on the time to maturity of the securities used to calibrate the model. Diebold [22], attempted to address the multi-colinearity problem by fixing the value of τ_s over time, which in some instances might not produce accurate results.

Due to the local-minima problem which makes model estimation difficult in the Nelson-Siegel [1] model and the inadequacy of the calibration methods used so far, we propose NLS estimation with L-BFGS-B method optimization approach. This optimization method is an extension of the limited memory BFGS method (LM-BFGS or L-BFGS) which uses simple boundaries model, according to Zhu *et al.* [32].

Using L-BGFS-B algorithm, we can estimate the above five parameters: $\varphi \equiv \{\beta_0, \beta_1, \beta_2, \tau_1, \tau_2\}$, embedded in the Nelson–Siegel [1] model, and hence calculate the price of the bond using the following nonlinear constrained optimization estimation procedure based the Gauss-Newton numerical method:

$$P_i = \sum_{m=1}^{T} \frac{CF_{im}}{\left\{1 + \beta_0 + \beta_1\left(\frac{\tau_1}{m}\right)\left[1 - \exp\left(\frac{-m}{\tau_1}\right)\right] + \beta_2\left(\frac{\tau_2}{m}\right)\left[1 - \exp\left(\frac{m}{\tau_2}\right)\left(\frac{m}{\tau_2} + 1\right)\right]\right\}^m} + \varepsilon_i \tag{6}$$

where P_i is the price of bond i.

3.1.2. Improvement of Monotone Convex Interpolation on $r(t)t$

We start with a mesh of data points $\{t_1, t_2, \cdots, t_n\}$ (we will think of the x-values as time points on the x axis) and the corresponding y values are define as $\{f_1, f_2, \cdots, f_n\}$ for a generic but unknown function $f(t)$. Cubic splines are generally defined by piece-wise cubic polynomial that passes through consecutive points:

$$f(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i + (t - t_i)^3. \tag{7}$$

With $t \in [t_i, t_{i+1}]$ and $i = 1, \cdots, n$. We will use the following definitions:

$$h_i = t_{i+1} - t_i \tag{8}$$

$$m_i = \frac{f_{i+1} - f_i}{h_i}. \tag{9}$$

With $i = 1, \cdots, (n-1)$ The coefficients a_i, b_i, c_i and d_i depends on the details of the method, and are related to the values of $f(t)$ and its derivatives at the node points. In general

$$a_i = f(t_i) \equiv f_i, \ b_i = f'(t_i) \ \text{and so on} \tag{10}$$

where in the above equation, the prime denotes the derivative of the interpolating function $f(t)$ w.r.t. its argument t. Moreover given a_i and b_i, we can express c_i and d_i as follows:

$$c_i = \frac{3m_i - b_{i+1} - 2b_i}{h_i} \tag{11}$$

$$d_i = \frac{b_{i+1} - b_i - 2m_i}{h_i^2} . \tag{12}$$

We can use Equation (4) to compute the following derivative:

$$\frac{\partial f(t)}{\partial f_j} = \frac{\partial a_i}{\partial f_j} = \frac{\partial b_i}{\partial f_j}(t - t_i) + \frac{\partial c_i}{\partial f_j}(t - t_i)^2 + \frac{\partial a_i}{\partial f_j}(t - t_i)^3 . \tag{13}$$

We have:

$$\frac{\partial a_i}{\partial f_j} = \delta_i^j \tag{14}$$

$$\frac{\partial m_i}{\partial f_j} = \frac{1}{h_i}\delta_{i+1}^j - \frac{1}{h_i}\delta_i^j \tag{15}$$

$$\frac{\partial c_i}{\partial f_j} = \frac{1}{h_i}\left(3\frac{\partial m_i}{\partial f_j} - \frac{\partial b_{i+1}}{\partial f_j} - 2\frac{\partial m_i}{\partial f_j}\right) \tag{16}$$

$$\frac{\partial d_i}{\partial f_j} = \frac{1}{h_i^2}\left(\frac{\partial b_{b+1}}{\partial f_j} - \frac{\partial b_i}{\partial f_j} - 2\frac{\partial m_i}{\partial f_j}\right). \tag{17}$$

Which depends on the matrix element $\frac{\partial b_i}{\partial f_j}$. Here δ_i^j is the Kronecker delta, which is equal to one if $i = j$ and zero otherwise. Once the derivatives at the points, or equivalently the b_i coefficients, are specified, everything else is fixed. In particular we are interested in computing $\frac{\partial f(t)}{\partial f_j}$, then all the work will be in the calculation of the derivatives b_i w.r.t. f_j.

This calculation is tricky if we use monotone preserving splines (or any other method which enforces monotonicity where the b_is are non-differentiable functions of the f_js (which involve the min and max functions). And this is where the novelty of this paper comes in.

Let us start by recalling the formulas for the b_i's in the monotone preserving cubic spline method as defined in the Hagan-West [14]. First of all, at the boundaries, we have:

$$b_i = 0, \ b_n = 0 \tag{18}$$

For the internal data, if the curve is not a monotone at t_i, *i.e.* $m_{i-1} \cdot m_i \leq 0$, then the boundaries become:

$$b_i = 0 \ \left(\text{if } m_{i-1} \cdot m_i \leq 0\right) \tag{19}$$

So that it will have a turning point there. Instead if the trend is a monotone at t_i, *i.e.* $m_{i-1} \cdot m_i > 0$, then one defines:

$$\beta_i = \frac{3m_{i-1} \cdot m_i}{\max\left(m_{i-1}, m_i\right) + 2\min\left(m_{i-1}, m_i\right)} . \tag{20}$$

And:

$$b_i = \begin{cases} \min\left(\max\left(0, \beta_i\right), 3\min\left(m_{i-1}, m_i\right)\right) & \text{if } m_{i-1} \cdot m_i > 0 \\ \min\left(\max\left(0, \beta_i\right), 3\min\left(m_{i-1}, m_i\right)\right) & \text{if } m_{i-1} \cdot m_i < 0 \end{cases} . \tag{21}$$

The former choice is made when the curve is increasing (positive slopes), the latter when decreasing (negative slopes). Equation (21) represents the monotonicity constraint introduced by Hyman and based on the Fritsch-

Butland algorithm.

We can now move to the monotonicity constrain. What we need first is:

$$\frac{\partial}{\partial f_j} \max\left(0, \beta_i\right) = \begin{cases} 0 & \text{if } \beta_i < 0 \Leftrightarrow m_{i-1}, m_i < 0 \\ \frac{\partial \beta_i}{\partial f_j} & \text{if } \beta_i > 0 \Leftrightarrow m_{i-1}, m_i > 0 \end{cases} \tag{22}$$

$$\frac{\partial}{\partial f_j} \min\left(0, \beta_i\right) = \begin{cases} 0 & \text{if } \beta_i > 0 \Leftrightarrow m_{i-1}, m_i > 0 \\ \frac{\partial \beta_i}{\partial f_j} & \text{if } \beta_i < 0 \Leftrightarrow m_{i-1}, m_i < 0 \end{cases} . \tag{23}$$

The final step consists of putting all the information together to compute $\frac{\partial b_i}{\partial f_j}$. Suppose first that the trend is

increasing, *i.e.* $m_{i-1}, m_i, \beta_i > 0$. Using Equation (23) we find:

$$\frac{\partial b_i}{\partial f_j} = \frac{\partial}{\partial f_j}\left[\min\left(\max\left(0, \beta_i\right), 3 * \min\left(m_{i-1}, m_i\right)\right)\right] \tag{24}$$

$$= \begin{cases} \frac{\partial}{\partial f_j} \max\left(0, \beta_i\right) & \text{if } \max\left(0, \beta_i\right) < 3 * \min\left(m_{i-1}, m_i\right) \\ 3 * \frac{\partial}{\partial f_j} \min\left(m_{i-1}, m_i\right) & \text{if } \max\left(0, \beta_i\right) > 3 * \min\left(m_{i-1}, m_i\right) \end{cases} . \tag{25}$$

Let us now suppose that the trend is decreasing instead, *i.e.* $m_{i-1}, m_i, \beta_i < 0$. By Equation (23) we have:

$$\frac{\partial b_i}{\partial f_j} = \frac{\partial}{\partial f_j}\left[\max\left(\min\left(0, \beta_i\right), 3 * \max\left(m_{i-1}, m_i\right)\right)\right] \tag{26}$$

$$= \begin{cases} \frac{\partial}{\partial f_j} \min\left(0, \beta_i\right) & \text{if } \min\left(0, \beta_i\right) > 3 * \max\left(m_{i-1}, m_i\right) \\ 3 \frac{\partial}{\partial f_j} \max\left(m_{i-1}, m_i\right) & \text{if } \min\left(0, \beta_i\right) < 3 * \max\left(m_{i-1}, m_i\right) \end{cases} . \tag{27}$$

It can be shown that for any *i* and *j*

$$\frac{\partial}{\partial f_j} \max\left(m_{i-1}, m_i\right) = \begin{cases} \frac{\partial m_{i-1}}{\partial f_j} & \text{if } m_{i-1} > m_i \\ \frac{\partial m_i}{\partial f_j} & \text{if } m_{i-1} < m_i \end{cases} \tag{28}$$

$$\frac{\partial}{\partial f_j} \min\left(m_{i-1}, m_i\right) = \begin{cases} \frac{\partial m_{i-1}}{\partial f_j} & \text{if } m_{i-1} < m_i \\ \frac{\partial m_i}{\partial f_j} & \text{if } m_{i-1} > m_i \end{cases} . \tag{29}$$

This formula solves our problem of non-differentiability found in the monotone preserving convex on $r(t)t$.

3.2. Liquidity-Weighted Function

In yield curve construction, errors are caused by two reasons: (a) curve fitting mistakes and (b) presence of liquidity premium. The errors due to curve fitting arise from the calculations and can be avoided. But the error due to the presence of liquidity premium is reflective of market conditions and one cannot ignore them. Since the reliability of the term structure estimation heavily depends on the precision of the market prices according to

Subramanian [33], liquid and illiquid securities are a heterogeneous class and including them both in the term structure estimation process poses problems. Illiquid bonds are traded at a premium to compensate for their undesirable attribute in terms of a low price. Assigning equal weights to both types of errors will give undue weight to the kind of error that creeps in due to curve fitting.

Subramanian [33] suggests a liquidity weighted objective function, which hypothesizes that a weighted error function (with weights based on liquidity) would lead to better estimation that equal weights to the squared errors of all securities. We therefore model the liquidity using a function with two factors: the volume of trade in a security and the number of trades in that security.

The weight of the ith security W_i is given by:

$$W_i = \frac{\left[\left(1-e^{\left(-\frac{v_i}{v_{\max}}\right)}\right)+\left(1-e^{\left(-\frac{n_i}{n_{\max}}\right)}\right)\right]}{\sum_i W_i} = \left(1-e^{\left(-\frac{v_i}{v_{\max}}\right)}\right)+\left(1-e^{\left(-\frac{n_i}{n_{\max}}\right)}\right) \tag{30}$$

$$W_i = \frac{\left[\tanh\left(-\frac{v_i}{v_{\max}}\right)+\tanh\left(-\frac{n_i}{n_{\max}}\right)\right]}{\sum_i W_i} = \tanh\left(-\frac{v_i}{v_{\max}}\right)+\tanh\left(-\frac{n_i}{n_{\max}}\right) \tag{31}$$

where v_i and n_i are the volume of trade and the number of trades in the ith security respectively, while $v_{\max}$ and $n_{\max}$ are the maximum number of trades among all the securities traded for the day respectively.

As given in the Equation (30) and Equation (31) above, it ensures that the weights of the relative liquid securities would not be significantly different from each other. For the illiquid securities, however the weights would fall quickly as liquidity decreased.

The final error-minimizing function, which should equal to zero, is given by:

$$\text{Min}\sum_{i=1}^{n} w_i \left(P_i - B_i\right)^2 = \text{Min}\sum_{i=1}^{n} w_i \varepsilon_i^2 = 0. \tag{32}$$

3.3. Test Statistics

In academic literature, there are two distinct approaches used to indicate the term structure fitting performance. One is the flexibility of the curve (accuracy), and the other focuses on smoothness of the yield curve. Although there are numerical methods proposed to estimate the term structure, any method developed has to grapple with deciding the extent of the above trade-off. Hence it becomes a crucial issue to investigate how to reach a compromise between the flexibility and smoothness.

Three simple summary statistics which can be calculated for the flexibility of the estimated yield curve are the coefficient of determination, root mean squared percentage error, and root mean squared error . These are calculated as shown in the next four subsections.

3.3.1. The Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(P_i - \hat{B}_i\right)^2 \Big/ (n-k)}{\sum_{i=1}^{n}\left(P_i - \hat{P}_i\right)^2 \Big/ (n-1)} \tag{33}$$

where $\overline{P}$ is the mean average price of all observed bonds, $\hat{B}_i$ is the model price of a bond i, n the number of bonds traded and k is the number of parameters needed to be estimated.

Roughly speaking, with the same analysis in regression, we associate a high value of R^2 with a good fit of the term structure and associate a low R^2 with a poor fit.

3.3.2. Root Mean Squared Error (RMSE)
Denoted as the RMSE, a low value for this measure is assumed to indicate that the model is flexible, on average, and is able to fit the yield curve.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(P_i - \hat{B}_i \right)^2} . \tag{34}$$

3.3.3. Root Mean Squared Percentage Error (RMSPE)

Denoted as the RMSPE, a low value for this measure is also assumed to indicate that the model is flexible, on average, and is able to fit the yield curve.

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{P_i - \hat{B}_i}{P_i} \right)^2} * 100\% . \tag{35}$$

3.3.4. Testing for Smoothness

To test the smoothness of the estimated yield curve, we use a modified statistic suggested by Adams and Deventer [34] to reach the maximum smoothness for forward rate curve, and denote the smoothness (Z) for the estimated yield curve as:

$$Z = \sum_{t=1.5}^{n} \left(\left[f(t) - f\left(t - \frac{1}{2} \right) \right] - \left[f\left(t - \frac{1}{2} \right) - f(t-1) \right] \right)^2 \times \frac{1}{2} = 0 . \tag{36}$$

Ideally, the value should equal to zero. The model with the least Z value is deemed to be the best.

4. Empirical Results

4.1. Data

In Kenya, nearly all bond transactions take place on the OTC market. The data used in this study was supplied by the Central Bank of Kenya. The sample period contains 417 weekly data from January 2005 to December 2012. Weekly prices (every Friday) for 108 Kenyan Government Bonds (KGBs) with original maturity dates ranged from 2 to 30 years are obtained.

4.2. Parameter Estimation

4.2.1. Nelson-Siegel [1] Model

Table 1 lists the summary statistics of estimated parameters for the Nelson-Siegel [1] model. It is seen that all estimated values for $\hat{\beta}_1$ and $\hat{\beta}_2$ are negative, which indicates that the yield curves generated by this model are all positively and upward sloping without a visible hump.

4.2.2. The Spline Model

Table 2 lists the summary statistics of estimated parameters for the improved monotone preserving interpolation method on $r(t)t$ Model.

4.3. Comparison of the Models

4.3.1. In terms of Accuracy

A direct comparison of the two models in Table 3 appears to favor the Nelson-Siegel [1].

4.3.2. In Terms of Smoothness

In the academic literature, it has been observed that when comparing alternative methods of term structure fitting models, there is usually a trade-off between flexibility and smoothness. In Table 4, the spline seems to have the best fit in flexibility for fitting the term structure of KGB market. However, as shown in Table 4, the Nelson-Siegel [1] Model is superior to the spline, which shows that the Nelson-Siegel [1] results to a relatively smoother yield curve, compared to the other model.

4.3.3. In Terms of the Spot Curves

In Figure 1, we see the spot curve generated by the Nelson and Siegel [1] model, which shows character-

Table 1. Results for estimated parameters for Nelson-Siegel [1] model.

Year	Parameters			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}_1$
2005	0.0587	−0.0115	−0.0040	4.6089
2006	0.0463	−0.0090	−0.0127	3.2148
2007	0.0454	−0.0133	−0.0388	1.8327
2008	0.0356	−0.0183	−0.0425	2.1674
2009	0.0358	−0.0164	−0.0493	1.0232
2010	0.0225	−0.0085	−0.0819	0.6237
2011	0.0225	−0.0085	−0.0819	0.6237
2012	0.0241	−0.091	−0.0213	1.0595

Table 2. Results for estimated parameters for improved monotone preserving interpolation method on $r(t)t$.

Year	parameters			
	a_i	b_i	c_i	d_i
2005	0.0588	−0.0125	0.0316	−0.0365
2006	0.0477	−0.0152	−0.0034	0.0010
2007	0.0400	−0.0284	0.0701	−0.0410
2008	0.0355	−0.0293	0.0070	−0.0189
2009	0.0359	−0.0308	−0.0068	0.0115
2010	0.0232	−0.0022	−0.0074	−0.0110
2011	0.0232	−0.0022	−0.0074	−0.0110
2012	0.0279	−0.0061	−0.0074	−0.0055

Table 3. Summary statistics for fitting performance in terms of accuracy.

	RMSPE		RMSE		R^2	
	Nelson-Siegel	Spline Method	Nelson-Siegel	Spline Method	Nelson-Siegel	Spline Method
mean	0.0144	0.0122	1.6318	1.4043	0.9654	0.9738
Std.dev	0.0066	0.0051	0.7752	0.6033	0.0357	0.0299
Max	0.0413	0.0281	4.6311	3.3047	0.9973	0.9979
Min	0.0050	0.0041	0.5311	0.4363	0.8015	0.8219

Table 4. Summary Statistics for fitting performance in terms of smoothness.

Model	With liquidity constraint	
	R^2	Smoothness: $(Z)(\times 10^{-6})$
Nelson-Siegel [1]	0.9654	4.9822
Spline (interpolation)	0.9738	10.7467

ristics of normal curve.

In **Figure 2**, we see that the splines produce a curve which initially is increasing and then suddenly becomes flat after time 10, at a constant yield rate of 1.40%. This does not depict normal behavior of yield curves.

4.3.4. In Terms of Pricing Back the Securities

In **Figure 3**, we see the Nelson and Siegel model [1] prices back the securities but not with as much accuracy as the spline curve does, as shown in **Figure 4**.
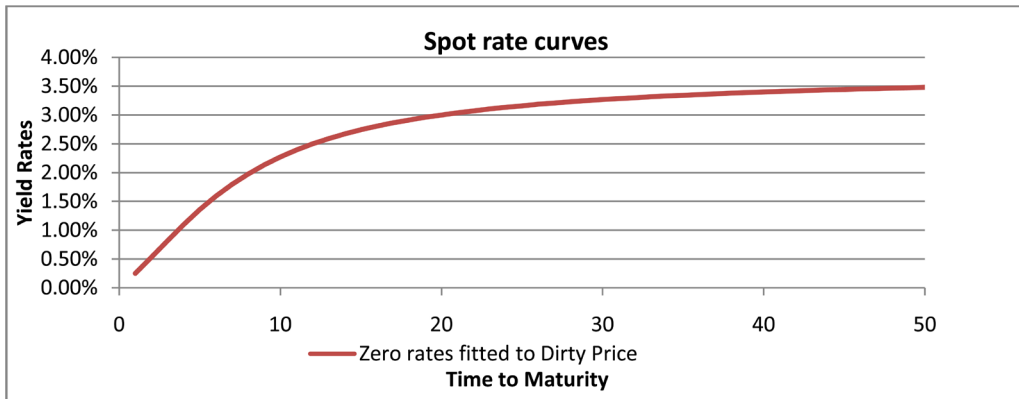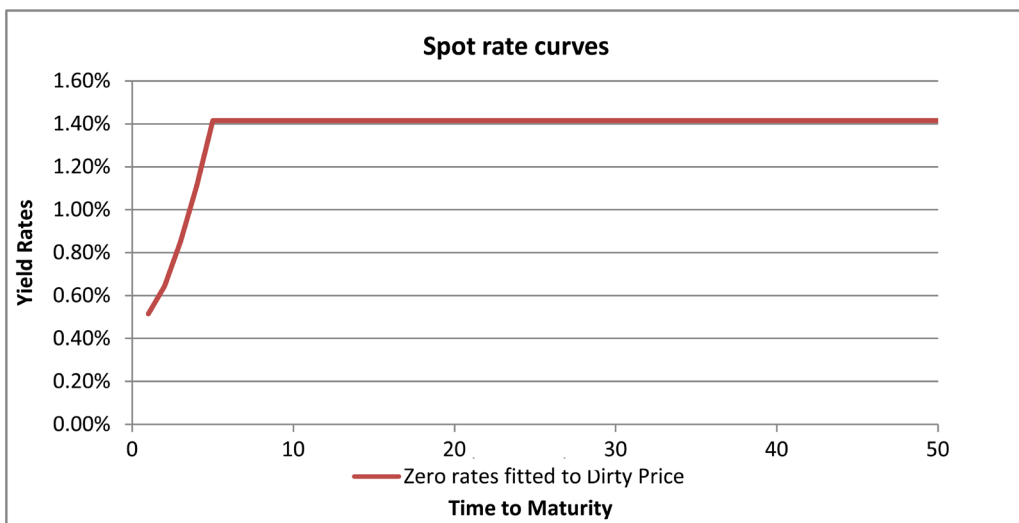
Figure 1. Nelson and siegel [1] curve.
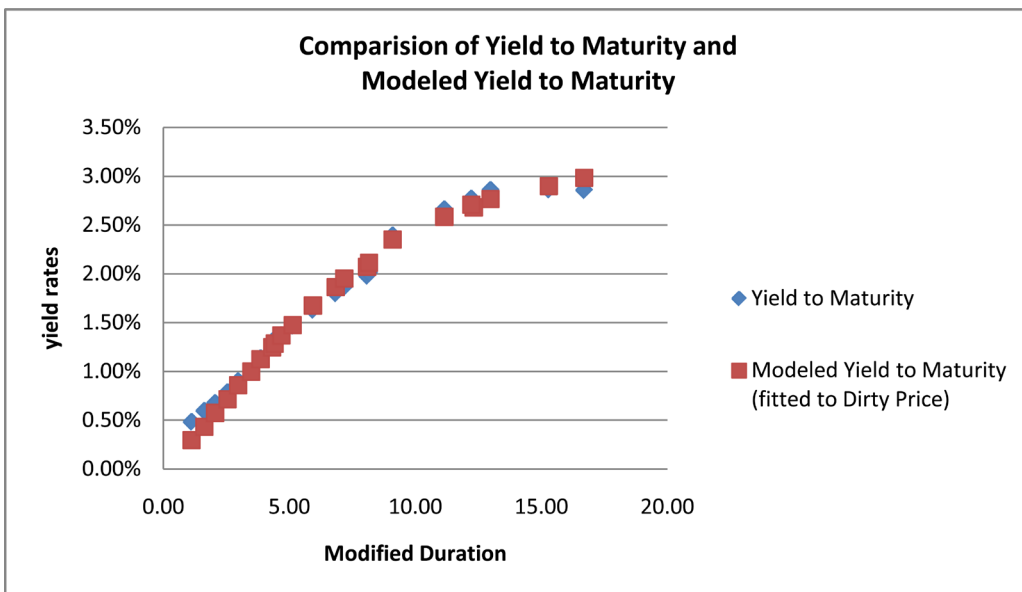

Figure 2. Splines curve.

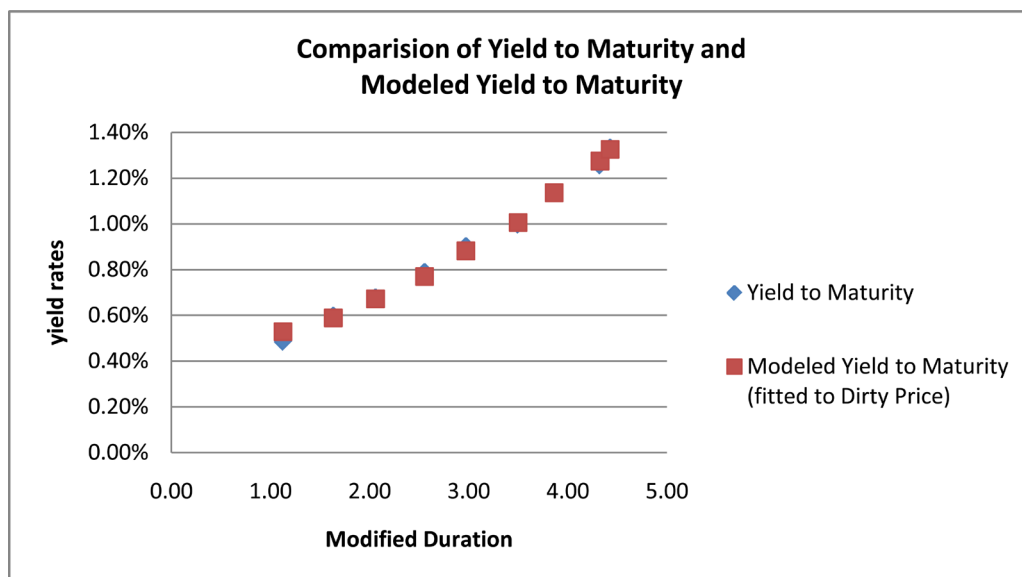**Figure 3.** Nelson and Siegel [1].
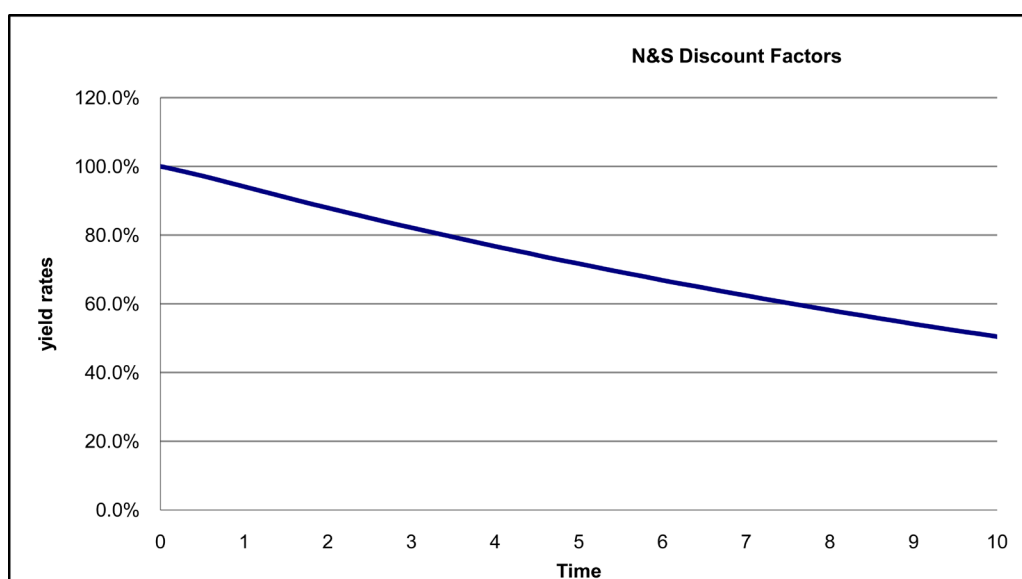
Figure 4. Splines.

Figure 5. The Nelson-Siegel model [1].

4.3.5. In terms of the Discount Curve

In **Figure 5**, we see that the Nelson and Siegel [1] model is able to produce a decreasing discount curve, which is a very important characteristic of a yield curve because it shows not only monotonicity, but is also sensible from an economic perspective. When we compare with **Figure 6**, we see that the discount curve is decreasing in the short to middle terms, with a little hitch at times 10 - 12, 16 - 28 and 29 - 31, after which it starts increasing. This curve lacks monotonicity and therefore does not make much economic sense.

5. Conclusion and Discussion of Results

The objective of this paper was to compare the performance of the improved monotone preserving interpolation method on $r(t)t$ against the Nelson & Siegel [1] in terms of several performance yard sticks. These yard sticks were: 1) accuracy, 2) smoothness, 3) the non-negativity and continuity of forward and spot curves, 4) pricing back of securities and 5) the monotonicity of the discount curve.
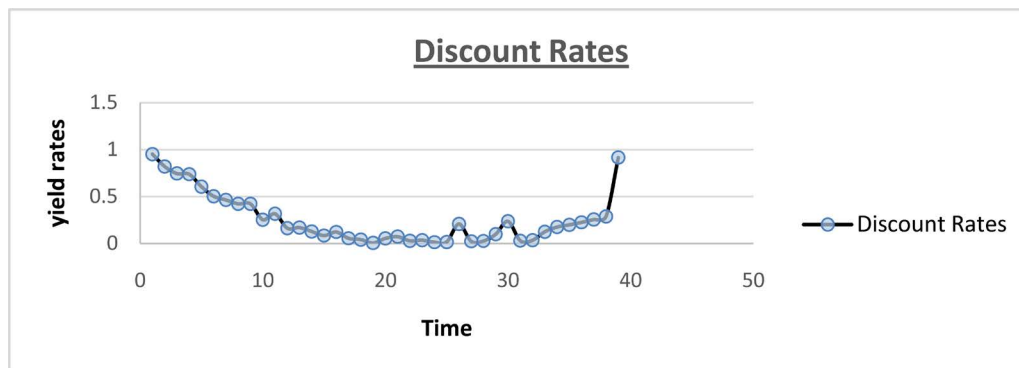
Figure 6. Splines discount curve.

In terms of accuracy, we see that the spline method does better than the parametric method in that it gives an accuracy of 97.37% compared to an accuracy of 96.54% by the Nelson-Siegel [1] model. We come to the same conclusion when we compare the two models using the test statistics of accuracy in that the spline does better than the parametric model.

However, when it comes to the test of smoothness, we see that Nelson-Siegel [1] model performs much better with a lower Z value of 4.9822 compared to the spline's value of 10.7467. Smoothness of the curve is extremely important in that it points towards differentiability and continuity of the curve.

When we compare the spot and the forward curves, we see that Nelson-Siegel gives us better curves (normal curves) which reflect reality, compared to the resulting splines' curve which increases with time initially, then abruptly becomes static at a given yield rate. This behavior is not realistic. We come to the same conclusion (that the parametric is better than the spline) when we compare the two in terms of pricing back the fixed income securities in that we see an almost perfect pricing back of the securities when we use the parametric model.

When we compare the two in terms of monotonicity, a very important concept in both Mathematical Finance and Economics, we see again that Nelson-Siegel [1] performs better than the spline. It results to a strictly decreasing discount curve, compared to the decreasing spline curve which starts increasing in the long term (after 25 years).

After considering all the factors at hand, we come to a conclusion that the parametric model (Nelson-Siegel [1]) is the proper model to be used in NSE, and consequently, in the East African securities Markets.

Acknowledgements

References

[1] Nelson, C.R. and Siegel, A.F. (1987) Parsimonious Modelling of Yield Curves. *Journal of Business*, **60**, 473-489. http://dx.doi.org/10.1086/296409

[2] Muthoni, L., Onyango, S. and Ongati, O. (2015) Construction of Nominal Yield Curve for Nairobi Securities Exchange: An Improvement on Monotone Preserving Interpolation Method on r(t)t. *Journal of Mathematical Finance*.

[3] Muthoni, L., Onyango, S. and Ongati, O. (2015) Extraction of Zero Coupon Yield Curve for Nairobi Securities Exchange: Finding the Best Parametric Model for East African Securities Markets. *SIAM Journal on Financial Mathematics*.

[4] Vasiceck, O. (1977) An Equilibrium Characterization of the Term Structure. *Journal of Financial Economies*, **5**,

177-188. http://dx.doi.org/10.1016/0304-405X(77)90016-2

[5]   Dothan, L. (1978) On the Term Structure of Interest Rates. *Journal of Financial Economies*, **6**, 59-69. http://dx.doi.org/10.1016/0304-405X(78)90020-X

[6]   Brennan, M.J. and Schwartz, E.S. (1979) A Continuous Time Approach to the Pricing of Bonds. *Journal of Banking and Finance*, **3**, 133-156. http://dx.doi.org/10.1016/0378-4266(79)90011-6

[7]   Cox, J.C., Ingersoll Jr., J.E. and Ross, S.A. (1985) A Theory of the Term Structure of Interest Rates. *Econometrica*, **53**, 385-408.

[8]   Duffie, D. and Kan, R. (1996) A Yield Factor Model of Interest Rates. *Mathematical Finance*, **6**, 379-406. http://dx.doi.org/10.1111/j.1467-9965.1996.tb00123.x

[9]   Hull, J. and White, A. (1990) Pricing Interest Rate Derivatives Securities. *Review of Financial Studies*, **3**, 573-592. http://dx.doi.org/10.1093/rfs/3.4.573

[10]   Heath, D., Jarrow, R. and Morton, A. (1992) Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contigent Claim Valuation. *Econometrica*, **60**, 77-105. http://dx.doi.org/10.2307/2951677

[11]   Alper, C.E., Akdemir, A. and Kazimov, K. (2004) Estimating the Term Structure of Government Securities in Turkey. Bogazaci University Research Papers: ISS/EC-2004-03. http://dx.doi.org/10.2139/ssrn.578921

[12]   de Boor, C. (1977) Package for Calculating with B-Splines. *SIAM Journal on Numerical Analysis*, **14**, 441-472. http://dx.doi.org/10.1137/0714026

[13]   de Boor, C. (1978) A Practical Guide to Splines. Springer-Verlag, New York. http://dx.doi.org/10.1007/978-1-4612-6333-3

[14]   West, H. (2006) Methods for Constructing a Yield Curve. *Applied Mathematical Finance*, **13**, 89-129. http://dx.doi.org/10.1080/13504860500396032

[15]   Anderson, B.L. (2007) The Demise of the Identity Hypothesis and the Insufficiency and Non-Necessity of Contour Relatability in Predicting Object Interpolation: Comment on Kellman, Garrigan, and Shipley (2005). *Psychological Review*, **114**, 470-487. http://dx.doi.org/10.1037/0033-295X.114.2.470

[16]   Preez, D. (2013) A New Method for Interpolating Yield Curve Data in a Manner That Ensures Positive and Continuous Forward Curve. *South African Journal of Economic and Management Sciences*.

[17]   de Boor, C. (2001) Calculation of Smoothing Spline with Weighted Roughness Measure. *Mathematical Models and Methods in Applied Sciences*, **11**, 33-41. http://dx.doi.org/10.1142/S0218202501000726

[18]   Carlson, F. (1980) Monotone Piecewise Cubic Interpolation. *SIAM Journal of Numerical Analysis*, **17**, 238-246. http://dx.doi.org/10.1137/0717021

[19]   Hyman, J.M. (1983) Accurate Monotonicity Preserving Cubic Interpolation. *SIAM Journal of Science and Statistical Computing*.

[20]   Akima, H. (1970) A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures. *Journal of ACM*, **17**, 589-602. http://dx.doi.org/10.1145/321607.321609

[21]   Chambers, C.W. (1984) A New Approach to Estimation of the Term Structure of Interest Rate. *Journal of Financial and Qualitative Analysis*, **19**, 233-252. http://dx.doi.org/10.2307/2331088

[22]   Diebold, L. (2006) Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics*, **130**, 337-364. http://dx.doi.org/10.1016/j.jeconom.2005.03.005

[23]   Modena, M. (2008) An Empirical Analysis of the Curvature Factor of the Term Structure of Interest Rates. MPRA Paper, No. 11597.

[24]   Tam, Y. (2008) Modelling Sovereign Bond Yields Curves of the US, Japan and Germany. *International Journal of Finance and Economics*, **13**, 82-91. http://dx.doi.org/10.1002/ijfe.353

[25]   Svensson, L.E.O. (1992) Estimating and Interpreting Forward Interest Rates. NBER Working Paper Series 4871.

[26]   Bliss, R.R. (1996) Testing Term Structure Estimation Methods. Working Paper 96-12a, Federal Reserve Bank of Atlanta.

[27]   Bolder, S. (1999) Yield Curve Modelling at the Bank of Canada. Technical Report, 84. http://dx.doi.org/10.2139/ssrn.1082845

[28]   Maria, L.M. (2009) Estimation of the Term Structure of Interest Rates: The Venezuelan Case. *Journal of Future Markets*, **23**, 1075-1105.

[29]   Gilli, G.S. (2010) Calibrating the Nelson-Siegel-Svensson Model. COMISEF Working Papers Series, 31. http://dx.doi.org/10.2139/ssrn.1676747

[30]   Rezende, F. (2011) Modelling and Forecasting the Yield Curve by an Extended Nelson and Siegel. Class of Models: A

Quatile Autoregression Approach. *Journal of Forecasting*, **32**, 111-123. http://dx.doi.org/10.1002/for.1256

[31] Rosadi, D. (2011) Modelling of Yield Curve and Its Computation by Software Packages RemdrPlugin. Econometric. *Proceeding FMIPA UnDip* 2011, 514-523.

[32] Zhu, C.Y., Byrd, R.H., Lu, P.H. and Nocedal, J. (1994) L-BFGS-B-Fortran Subroutines for Large-Scale Bound Constrained Optimization. Department of Electrical Engineering and Computer Science, Northwestern University, Evanston.

[33] Subramanian, K. (2001) Term Structure Estimation in Illiquid Markets. *The Journal of Fixed Income*, **11**, 77-86. http://dx.doi.org/10.3905/jfi.2001.319292

[34] Deventer, A. (1994) Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness. *Journal of Fixed Income*, **4**, 52-62. http://dx.doi.org/10.3905/jfi.1994.408102

[35] Broyden, C. (1970) The Convergence of a Class of Double-Rank Minimization Algorithms. II. The New Algorithm. *Journal of the Institute of Mathematics and Its Applications*, **6**, 222-231. http://dx.doi.org/10.1093/imamat/6.3.222

[36] Fletcher, R. (1970) A New Approach to Viable Metric Algorithms. *The Computer Journals*, **13**, 317-322. http://dx.doi.org/10.1093/comjnl/13.3.317

[37] Goldfarb, D. (1970) A Family of Variable Metric Methods Derived by Variational Means. *Mathematics of Computation*, **24**, 23-26. http://dx.doi.org/10.1090/S0025-5718-1970-0258249-6

[38] Shanno, D. (1970) Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, **24**, 647-656. http://dx.doi.org/10.1090/S0025-5718-1970-0274029-X

[39] Nocedal, J. (1980) Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation*, **35**, 773-782. http://dx.doi.org/10.1090/S0025-5718-1980-0572855-7

[40] Nocedal, J. and Wright, S.J. (1999) Numerical Optimization. Springer Series in Operations Research, Springer-Verlag, New York.

[41] Lewis, A.S. and Overton, M.L. (2013) Nonsmooth Optimization via Quasi-Newton Methods. *Mathematical Programming*, **141**, 135-163. http://dx.doi.org/10.1007/s10107-012-0514-2

[42] Conn, A.R., Gould, N.I.M. and Toint, P.L. (1988) Global Convergence of a Class of Trust Region Algorithms for Optimization with Simple Bounds. *SIAM Journal on Numerical Analysis*, **25**, 433-460. http://dx.doi.org/10.1137/0725029

[43] Moré, J.J. and Toraldo, G. (1989) Algorithms for Bound Constrained Quadratic Programming Problem. *Numerische Mathematik*, **55**, 377-400. http://dx.doi.org/10.1007/BF01396045

Appendix 1: The L-BFGS-B Algorithm

A.1. Introduction

The problem addressed is to find a local minimizer of the non-smooth minimization problem.

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^n} f(x) \\
&s.t. \;\; l_i \leq x_i \leq u_i \\
&i = 1, \cdots, n.
\end{aligned}
\tag{A1}
$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuous but not differentiable anywhere and n is large. l_i and u_i are respectively an upper limit and; lower limit parameters. $f(x)$ is NLS (Non Linear Schrödinger) function of residual functions of Nelson-Siegel [1] model class and x is a parameter of the Nelson-Siegel [1] model class.

The L-BFGS-B algorithm by Zhu *et al.* [32] is a standard method for solving large instances of $\min_{x \in \mathbb{R}^n} f(x)$ when f is a smooth function, typically twice differentiable. The name BFGS stands for Broyden, Fletcher, and Goldfarb and Shanno, the originators of the BFGS quasi-Newton algorithm for unconstrained optimization discovered and published independently by them in 1970 [Broyden [35], Fletcher [36], Goldfarb [37] and Shanno [38]. This method requires storing and updating a matrix which approximates the inverse of the Hessian $\nabla^2 f(x)$ and hence requires $\mathcal{O}(n^2)$ operations per iteration.

According to Nocedal [39], the L-BFGS variant where the L stands for "Limited-Memory" and also for "Large" problems, is based on BFGS but requires only $\mathcal{O}(n)$ operations per iteration, and less memory. Instead of storing the $n \times n$ Hessian approximations, L-BFGS stores only m vectors of dimesion n, where m is a number much smaller than n. Finally, the last letter B in L-BFGS stands for bounds, meaning the lower and upper bounds l_i and u_i. The L-BFGS-B algorithm is implemented in a FORTRAN software package, according to by Zhu *et al.* [32]. We discuss how to modify the algorithm for non-smooth functions.

A.1.1. BFGS

BFGS is standard tool for optimization of smooth functions. It is a line search method. The search direction is of type $d = -B_k \nabla f(x_k)$ where B_k approximation to the inverse Hessian of f.[2] This k^{th} step approximation is calculated via the BFGS formula.

$$
B_{k+1} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} \right) B_k \left(I - \frac{y_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} \right) + \frac{s_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}
\tag{A2}
$$

where $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ and $s_k = x_{k+1} - x_k$. BFGS exhibits super-linear convergence on generic problems but it requires $\mathcal{O}(n^2)$ operations per iteration, according to Wright *et al.* [40].

In the case of non-smooth functions, BFGS typically succeeds in finding a local minimizer, as indicated by Overton *et al.* [41]. However, this requires some attention to the line search conditions. This conditions are known as the Armijo and weak Wolfe line search conditions and they are a set of inequalities used for computation of an appropriate step length that reduces the objective function" sufficiently"

A.1.2. L-BFGS

L-BFGS stands for Limited-memory BFGS. This algorithm approximates BFGS using only a limited amount of computer memory to update an approximation to the inverse of the Hessian of f. Instead of storing a dense $n \times n$ matrix, L-BFGS keeps a record of the last m is a small number that is chosen in advance. For this reason the first m iterations of BFGS and L-BFGS produce exactly the same search directions if the initial approximation of B_0 is set to the identity matrix.

Due to this construction, the L-BFGS algorithm is less computationally intensive and requires only $\mathcal{O}(mn)$ operations per iteration. So it is much better suited for problems where the number of dimensions n is large.

A.1.3. L-BFGS-B

Finally L-BFGS-B is an extension of L-BFGS. The B stands for the inclusion of Boundaries. L-BFGS-B requires two extra steps on top of L-BFGS. First, there is a step called gradient projection that reduces the dimensionality of the problem. Depending on the problem, the gradient projection could potentially save a lot of iterations by eliminating those variables that

---

[2]When it is exactly the inverse Hessian this method is known as Newton method. Newton's method has quadratic convergence but requires the explicit calculation of the Hessian at every step.

are on their bounds at the optimum reducing the initial dimensionality of the problem and the number of iterations and running time. After this, gradient projection comes to second step of subspace minimization. During the subspace minimization phase, an approximate quadratic model of (A1) is solved iteratively in a similar way that the original L-BFGS algorithm is solved. The only difference is that the step length is restricted as much as necessary in order to remain within the lu-box defined by Equation (A1).

A.1.4. Gradient Projection

The L-BFGS-B algorithm was designed for the case when n is large and f is smooth. Its first step is the gradient projection similar to the one outlined in Conn, Gould, and Toint [42] and Toraldo, Jorge and Gerardo [43], which is used to determine an active set corresponding to those variables that are on either their lower or upper bounds. The active set is defined at point x^* is:

$$\mathcal{A}\left(x^*\right) = \left\{i \in \{1,\cdots,n\} \setminus x_i^* = l_i \vee x_i^* = u_i\right\} \tag{A3}$$

Working with this active set is more efficient in large scale problems. A pure line search algorithm would have to choose to step length short enough to remain within the box defined by l_i and u_i. So if at the optimum, a large number $\mathcal{B}$ of variables are either on the lower or upper bound, as many as $\mathcal{B}$ of iterations might be needed. Gradient projection tries to reduce this number of iterations. In the best case, only one iteration is needed instead of $\mathcal{B}$.

Gradient projections works on the linear part of the approximation model:

$$m_k\left(x\right) = f\left(x_k\right) + \nabla f\left(x_k\right)^{\mathrm{T}}\left(x - x_k\right) + \frac{\left(x - x_k\right)^{\mathrm{T}} H_k\left(x - x_k\right)}{2} \tag{A4}$$

where H_k is a L-BFGS-B approximation to the Hessian $\nabla^2 f$ stored in the implicit way defined by L-BFGS.

In this first stage of the algorithm a piece-wise linear path starts at the current point x_k in the direction $-\nabla f\left(x_k\right)$. Whenever this direction encounters one of the constraints the path runs corners in order to remain feasible. The path is nothing but feasible piece-wise projection of the negative gradient direction on the constraint box determined by the values l and u. At the end of this stage, the value of x that minimizes $m_k\left(x\right)$ restricted to this piece-wise gradient path is known as the "Cauchy point" x^c.

From this description of the estimation and optimization, following steps can be summarized:

- Find the residual function (r) of each model.

- Find NLS estimation, *i.e.* $f\left(x_i\right) = \frac{1}{2}\sum_{i=1}^{p}\left[x_i\right]^2$, of each model.

- Find the $p \times p$ matrix value for $B_1 = I$, p is the number of parameters estimated in each model.
- Find the initial value of parameter vector with rank $p \times 1$, p is the number of parameters estimated in each model.
- Find gradient from step 2 with every parameter in models. e.g. $\nabla f\left(x_i\right)_i$.
- Substitute the initial value of the parameter (step 3) to gradient of step 5 with result. e.g. $\nabla f\left(x_1\right)$.
- Find the value of p_1

Then we find the value of $f\left(x_1\right)$ so it will obtain of d_1 and s_1.