Comparing Term Structure Estimation Techniques: An Exercise with Brazilian Data

Matheus Stivali[1], José Augusto Fiorucci[2], and Raul Yukihiro Matsushita[2]

[1]Institute for Applied Economic Research

[2]Department of Statistics, University of Brasilia

XXIV Brazilian Finance Meeting
Curitiba, July 4-6 2024

Abstract

This text evaluates the empirical models of the Term Structure of Interest Rates, comparing the resulting estimates regarding goodness-of-fit, robustness to outliers, and smoothness. In addition to the descriptive statistics on these metrics, the Friedman test and the multiple comparison procedure were used to assess the statistical significance of differences among the models. Literature usually considers nonparametric or spline models in addition to the parsimonious function models, derived from Nelson and Siegel (1987)'s seminal work. We expand this set of models by considering LOESS and two Kernel regression specifications. For the evaluation, we used data from Brazilian interest rate derivatives over 1313 working days. Considering the surveyed literature, applying LOESS and Kernel regression and the use of multiple comparison procedure in the context of yield curve estimation are novel contributions.

JEL codes: E43, C14.

Keywords: Term Structure of Interest Rates, Yield Curve, interpolation, splines, Nelson-Siegel model, multiple comparison test.

1 Introduction

The term structure of interest rates, which depicts the relationship between time to maturity and the level of interest rates paid by a bond, is a fundamental concept with extensive applications in finance and macroeconomics. Whether pricing fixed-income debt instruments or extracting insights into inflation and activity level expectations embedded in the yield curve, the term structure plays a pivotal role.

Most uses of the term structure depend on it being fully observable, which does not happen. Therefore, estimating term structure using the available data points is necessary. The yield curve is expected to exhibit smoothness as it transitions between known data points and extrapolates beyond the last known maturity to estimate long-term interest rates. A yield curve estimate should ideally exhibit key economic characteristics, including non-negativity of interest rates and an upper limit on long-term rates.

The goal of this text is to compare various parametric and nonparametric techniques using recent Brazilian data, akin to the approaches taken by Ioannides (2003) for British data and Nymand-Andersen (2018) for European data. Previous studies for Brazil, such as those by Varga (2009) and Caldeira (2011), are approximately a decade old. Brazil's fiscal and monetary developments over this decade suggest that an up-to-date evaluation may provide better guidance. Additionally, these previous studies encompass different models than those examined here.

The estimation techniques are implemented using Brazilian data from January 2018 to April 2023. This timeframe encompasses diverse economic scenarios that give rise to the various yield curve forms reported in the literature. The comparison of these techniques considers their goodness-of-fit, robustness to outliers, and smoothness. The choice of an estimation technique may vary depending on the application's specific requirements. For instance, derivatives pricing may necessitate capturing even minor fluctuations in term structures, while macroeconomists may prefer a smoother curve, primarily concerned with its overall shape.

Three interrelated concepts – yield-to-maturity, discount function, and forward rates – allow us to describe the term structure. Since the data used in this text is already in a yield-to-maturity format, this concept forms the foundation for all modelling exercises.

This text is structured into five sections. The following section presents spline and function-based techniques, referring to their use in the yield curve estimation literature. It also presents the Kernel regression and LOESS models, which, to our knowledge, have not been applied to yield curve estimation so far. The third section introduces the criteria employed to compare the models, considering dimensions like in- and out-of-sample goodness-of-fit, robustness, and smoothness. Furthermore, it presents formal tests to compare model performance. Section four reports the data employed and outlines the estimation procedures for implementing the presented models. Section five unveils the results, and the final section concludes this text.

2 Background and Related Literature

Estimating the yield curve entails interpolating between known maturities and extrapolating the yield level beyond the final maturity. Since the seminal work of McCulloch (1971, 1975), the nonparametric approach has employed splines as a means to approximate the yield curve.[1] Following a different approach, a branch of *parsimonious* parametric models originated with the Nelson-Siegel model (Nelson and Siegel, 1987). These parametric models have undergone further refinement, resulting in occasionally more intricate specifications.

This section presents the statistical formulation of those models and the estimation strategies used in their implementation. Furthermore, we also consider how nonparametric regression (kernel and local regression) performs in the yield curve estimation. While nonparametric regression is not widely used in this context, there is no reason for not considering these models since their goal is aligned with the yield curve estimation.

2.1 Spline based models

A **spline** is a **piecewise polynomial**, a function formed by joining different – yet of a specific degree – polynomials at fixed points of its domain, the *knots*.

The sequence of points $\xi = \{k_0, k_1, \dots, k_K\}$, which partitions a given interval $[a, b] \subset \mathbb{R}$ into subintervals is called a *knots sequence* (or a *knot set*), where $a = k_0 < k_1 < \dots < k_K = b$. The points $k_1, \dots, k_{K-1}$ are called *interior knots*, and the points k_0 and k_K *boundary knots*, (Hämmerlin and Hoffman, 1991, p. 229-230). The spline corresponds to the polynomials fitted to the observations between two knots in each segment. And the smooth connection between different segments is ensured by imposing on the polynomials certain differentiability conditions.

A first approach to model a nonlinear relationship can be to approximate it by a set of differ-

[1] It is worth noting that McCulloch originally estimated the discount curve, and depending on the dataset, it may be more tractable to estimate either the discount or forward curves, with the yield-to-maturity curve subsequently derived from these alternatives.

ent linear relationships over different ranges of the independent variable. Hence, what is known as **piecewise linear regression**. Firstly in an *ad hoc* manner, one can set breakpoints (*knots*), over the predictor's relevant range, positioning them where the function appears to change inclination. After that, it is possible to define indicator functions from these knots.

Let say that over $\mathbb{R}_0^+$ one can identify that the yield curve increases from 0 to k_1, changes its inclination at k_1 and flattens after k_3. From the knots definition, it is possible to create the variables: $\mathbb{1}_{[k_1,k_2[}(m)$ and $\mathbb{1}_{[k_2,\infty[}(m)$. Then the following regression can be estimated:

$$y_t(m) = \beta_0 + \beta_1 m + \beta_2(m - k_1)\mathbb{1}_{[k_1,k_2[}(m) + \beta_3(m - k_2)\mathbb{1}_{[k_2,\infty[}(m) + \epsilon_{t,m}. \tag{1}$$

A more direct approach to model a nonlinear relationship is to use a **polynomial regression** since a high-degree polynomial can produce complex nonlinear relationships[2].

According to Bolder and Gusba (2002, 3-4), considering N+1 distinct points $\{(x_0, f(x_0)), (x_1, f(x_1)), \ldots, (x_{N+1}, f(x_{N+1}))\}$ and defining $\mathcal{P}_N$ as the set of all polynomials of degree at most N, there is a unique polynomial $p \in \mathcal{P}_N$ such that $p(x_i) = f(x_i)$. Thus one can fit $N + 1$ points uniquely with a polynomial of degree N.

Such polynomial regression of degree N can be written as equation (2) and estimated by least squares as an usual linear model, (James et al., 2021, p. 290):

$$y_t(m) = \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 m^3 + \ldots + \beta_N m^N + \epsilon_{t,m}. \tag{2}$$

Determining the coefficients β by least squares depends on inverting the *Vandermonde matrix*, which is plagued by numerical problems (Bolder and Gusba, 2002, p. 4-6; Howard, 2017, p. 97-99). Besides the computational problems, a higher degree polynomial regression has other issues. For example, although the curve will lie close to the sample points, it may fluctuate remarkably between them, which is known as *Runge's phenomenon*. Additionally, the curve can become too flexible and assume unexpected forms near the independent variable (m) boundaries. For these reasons, usually, one considers a polynomial of degree 3 or 4 at the maximum.

Approaching the term structure modelling problem from the discount curve perspective, McCulloch (1971, p.28-29) pointed out that because market participants give more weight to minor yield differences in the near future than in the far future, the term structure will have a more complex form at shorter maturities than at the long end. Thus, a low-degree polynomial will fit the data at longer

[2]This approach follows from the *Weierstrass Approximation Theorem* that states that for a given degree of error ϵ any continuous function can be approximated by a polynomial with degree higher enough, Estep (2002, p. 509-510).

maturities but will not conform with observations at shorter maturities. According to McCulloch, only an "*extremely high-order polynomial*" would fit both the long and short ends.

Chambers et al. (1984) assumed that term structure in its yield-to-maturity form can be expressed as a polynomial $y_t(m) = \sum_{n=1}^{N} \beta_{t,n} m^{n-1}$, like equation (2) above. However, this authors could not estimate this specification directly because he did not observe the interest rates in the yield-to-maturity form. His problem was similar to McCulloch's, his data originated from coupon bond prices, and he needed to estimate discount functions for coupon payments. Considering continuously compounded interest rates, his assumption in terms of discount function becomes $d_t(m) = e^{-y_t(m)\times m} = e^{-\sum_{n=1}^{N} \beta_{t,n} m^{n-1}}$. Then he estimated $P_t = \sum_{t=1}^{T} c_t e^{-\sum_{n=1}^{N} \beta_{t,n} (T-t)^{n-1}}$, where P_t and c_t are the bond price and the coupon payment at moment t respectively, by nonlinear least squares.

Chambers et al. consider polynomials from degree one to five and conclude that third or fourth-degree polynomials explain much of the variation. Their residual analysis supports McCulloch's statement about the curve's short-end complexity since they observe a lack of fit at shorter maturities. Even though his data complicated their estimation procedure, if they had yield-to-maturity data, they could have adopted a specification as equation (2) and used ordinary least squares.

The approaches presented above are special cases of the **linear basis expansion**. The *basis* corresponds to the original predictors (the $\mathbf{X}$ matrix), whereas its *expansion* corresponds to transformations of the original x_i variables by the *basis functions* $b_1(.)$, $b_2(.)$, ..., $b_K(.)$. These functions are known and fixed and their use result in more predictors (i.e. the original $\mathbf{X}$ matrix augmented with new columns created by the transformations).

For the linear interpolation, the basis function is the indicator function, $b_j(m_i) = \mathbb{1}_{[k_j,k_{j+1}[}(m_i)$, while for the polynomial regression, the basis function corresponds to $b_j(m_i) = m_i^j$. Using linear basis expansions increases the dimensionality of the predictor's matrix ($\mathbf{X}$), producing a more flexible regression model, (Berk, 2016, p. 42). Therefore, instead of fitting the model on the original predictors one estimates a linear model of y_i against $b_1(m_i)$, $b_2(m_i)$, ..., $b_K(m_i)$:

$$y_i = \beta_0 + \beta_1 b_1(m_i) + \beta_2 b_2(m_i) + \beta_3 b_2(m_i) + \ldots + \beta_K b_K(m_i) + \epsilon_i. \tag{3}$$

Splines are constructed defining basis functions that represent the polynomial and the knots over the predictor's relevant range. A common choice for the basis functions is considering **cubic splines** (i.e. a spline of degree 3 and order 4). As in linear interpolation, K knots divide the relevant predictor's range. Then, one can fit a polynomial function, as equation (4), using least squares for each data subset delimited by the knots:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i \qquad x_i \in [k_j, k_{j+1}[. \tag{4}$$

However, the cubic spline imposes some constraints on the polynomials to obtain a smooth function: continuity, and continuity of the first and second derivatives. Continuity means that the function should assume the same value when approaching a knot from either side. The same should happen when evaluating the first and second derivatives[3].

One way of including those restrictions into equation (4) is to add to the basis of a cubic polynomial ($b_1(m) = 1$, $b_2(m) = m$, $b_3(m) = m^2$, $b_4(m) = m^3$) a *truncated power basis* with a function for each knot, (James et al., 2021, p. 297). The truncated power basis functions have the form:

$$(m-k)_+^r \equiv [max\{m-k, 0\}]^r \qquad r = 1, 2, 3, \ldots \qquad \text{or alternatively}$$

$$b(m_i, k_j) = (m-k)_+^r = (m_i - k_j)^r \mathbb{1}_{>0}(m_i - k_j) = \begin{cases} (m_i - k_j)^r & \text{if } m_i > k_j \\ 0 & \text{if } m_i \leq k_j \end{cases}.$$

Considering the definition above, the cubic spline regression equation for K knots can be written as:

$$y_i = \beta_0 + \beta_1 m_i + \beta_2 m_i^2 + \beta_3 m_i^3 + \sum_{j=1}^{K} \theta_j (m-k)_+^r + \epsilon_i \tag{5}$$

$$= \beta_0 b_1(X) + \beta_1 b_2(X) + \beta_2 b_3(X) + \beta_3 b_4(X) + \sum_{j=1}^{K} \theta_j b(m_i, k_j) + \epsilon_i. \tag{6}$$

Hence, a basis representing a cubic spline with three knots ($\xi = \{k_1, k_2, k_3\}$, with $k_1 < k_2 < k_3$) would be composed of seven basis functions: $b_1(X) = 1$, $b_2(X) = X$, $b_3(X) = X^2$, $b_4(X) = X^3$, $b_{k_1}(X) = (m-k_1)_+^3$, $b_{k_2}(X) = (m-k_2)_+^3$, $b_{k_3}(X) = (m-k_3)_+^3$. Equation (5) can be estimated by least squares. Considering $K = 3$, three knots divide the predictor's range into four subsets.

Similarly to the polynomial regression, estimated values beyond the predictors' boundaries may behave unexpectedly. For instance, considering the yield curve, this behaviour could be a steep inclination (positive or negative) after the final maturity observed. This lack of stability at the end of the maturity range is corrected by imposing an additional constraint: linearity after the last observation. When the model includes the linearity constraint (second derivative equals zero at the terminal points, $f''(x_1) = f''(x_N) = 0$), it is called **natural cubic spline**. As pointed out by Boor (2001, 44), this restriction is somewhat arbitrary and produces increased errors near the ends.

[3]There are some alternatives, for instance, the *cubic Hermite* and the *cubic Bessel*, whose constraint is continuity of the first derivative, (Boor, 2001, 39-41).

The works McCulloch (1971, 1975) pioneered the use of splines for estimating the discount curve and, from this function, obtaining the yield curve and forward interest rates. While McCulloch (1971) employed quadratic splines, McCulloch (1975) used natural cubic splines. Because the interest rates considered in the term structure correspond to the zero-coupon bonds yield-to-maturity, and this kind of bond is scarce in most maturities, McCulloch was initially concerned with extracting information from coupon bonds, i.e. bonds that regularly pay interest before maturity, equivalent to that from zero-coupon bonds. For that, he considered a coupon bond as a portfolio in which each coupon payment is a zero-coupon bond. In this perspective, the discount function estimate was the key to disentangle each synthetic zero-coupon bond from the original coupon bond price.

McCulloch (1971) expressed the problem of choosing the appropriate number of knots and their positions regarding the bias-variance trade-off[4] even though he did not use this terminology. McCulloch suggested that one should choose the number of knots to minimise the unbiased estimator of the variance. However, because this is a computationally expensive method with more than one local minimum, this author suggests determining the number of knots (K) as a function of the sample size (n) considering the closest integer to $K(n) = n^{1/2}$. Then the knots should be placed such that each segment has a similar number of observations (i.e. rounding $n/K(n)$).

An essential feature of the methods discussed so far is that the quantity and position of the knots are determined before the estimation. This is a characteristic of **regression splines**, and one can either follow McCulloch's rule or define the number and placement of the knots by comparing different alternatives as a model selection problem. However, this is cumbersome due to many possible candidates, (Berk, 2016, 64-65). The number and position of knots directly influence the final estimate, with smoothness being a function of the number of interior knots for the cubic splines. **Smoothing splines** present an alternative approach that incorporates the determination of the knot set into the estimation process. This can be achieved using the maximal set of knots and penalised regression.

To restate the problem, our objective is to achieve the best possible approximation of the underlying unknown function $y_t(m)$ using a spline $f_t(m)$:

$$y_t(m_i) = f_t(m_i) + \epsilon_{t,i} \qquad \epsilon_{t,i} \overset{iid}{\sim} N(0, \sigma^2).$$

However, including all data points as knots and simply minimising $RSS = \sum_{i=1}^{N}(y_i - f_t(m_i))^2$ will interpolate all data points, and $f_t(m)$ will not be smooth. Therefore, regularisation is necessary

[4]"*If k is too low, we will not be able to fit the discount function closely when it takes on difficult shapes. If it is too high, the discount function may conform too closely to outliers instead of being smooth. If k is as high as n, there will be no way to estimate σ^2.*", McCulloch (1971, p. 31).

to balance fit quality and smoothness. This can be done considering equation (7), in which $\lambda \geq 0$ is a *tuning parameter*

$$RSS(f_t, \lambda_t) = \sum_{i=1}^{N} [y_i - f_t(m_i)]^2 + \lambda_t \int [f_t''(m_i)]^2 \, dt. \tag{7}$$

While including a knot will improve the fit, reducing the first term on the right-hand side, it will also increase the penalty term, the second term on the right-hand side. The penalty term is the sum of the $f_t(.)$ second derivative at each data point and corresponds to a measure of the roughness of $f_t(.)$.

James et al. (2021, p. 302) point out that the minimisation of equation (7) results in a shrunken version of the natural cubic spline. The tuning parameter λ_t adjusts the bias-variance trade-off. If $\lambda_t = 0$, the penalty term will not affect the estimation, interpolating all points. On the other hand, a high value for λ_t will result in an ordinary linear regression. Therefore smoothing splines replace the knots' set choice by the tuning parameter determination.

One way to determine λ_t using the data is to minimise the cross-validated RSS, James et al. (2021, p. 302). This option can balance bias and variance on the final estimate, (Berk, 2016, p. 83-84).

Fisher et al. (1995) applied the smoothing splines technique to model the discount curve and forward rate function, considering cubic B-spline basis. Furthermore, they considered generalised cross-validation (GCV)[5] to determine the appropriate tuning parameter. To evaluate the performance, they proceeded with Monte Carlo simulations, simulating the term structure of the interest rates and then interpolating the data with smoothing splines; secondly, they applied it to seven years of US data. Trying to recover the "true" parameters from the simulated term structure the authors report that the selection of λ by GCV resulted in the least biased and most accurate estimates. However, when considering actual data this procedure was not so accurate but the estimates were coherent with financial theory.

Bliss (1996) noticed that the Fisher et al. (1995) method tends to have a worse performance at the short end. As highlighted before, the term structure usually has a more complex shape at the short end. Furthermore, the short end tends to be more populated by data points than the long end.

Waggoner (1997) proposed a way to handle these characteristics within the smoothing splines framework. Since the problem with the yield curve demands more flexibility at the short end than at the long end, this author proposed a variable tuning parameter $\lambda(m)$. Instead of selecting the tuning

[5] Cross-validation is a methodology that entails partitioning the given dataset into training and validation sets to assess the model's performance in an out-of-sample context, specifically regarding prediction error. Generalized cross-validation approximates the "leave-one-out" cross-validation approach, furnishing a criterion to appraise model performance by striking a balance between data fitting and model complexity.

parameter by GCV, it follows an *ad hoc* step function of maturity that is small for shorter maturities and large for long maturities. Because Waggoner was concerned with the US term structure, he defined the steps at maturities associated with different bonds traded in that market (bills, notes, and bonds). This specification is shown in equation (9); notice that maturity is measured in years in this case:

$$RSS(f,\lambda) = \sum_{i=1}^{M} [y_i - f(m_i)]^2 + \lambda(m) \int [f''(m)]^2 \, dt \tag{8}$$

$$\lambda(m) = \begin{cases} 0.1 & 0 \leq m \leq 1 \\ 100 & 1 \leq m \leq 10 \cdot \\ 100.000 & 0 \leq 10 \leq m \end{cases} \tag{9}$$

To evaluate his method, Waggoner (1997) sets his results against those obtained with McCulloch and Fisher et al. techniques. He considered two measures of goodness-of-fit (weighted mean absolute error and *hit rate*) and a smoothness measure. Considering in-sample goodness-of-fit, the variable tuning parameter method at shorter maturities performed better than the regular smoothing splines (Fisher et al. specification) and slightly better than McCulloch's natural cubic spline. On the other hand, Fisher et al. specification performed better at longer maturities.

Anderson and Sleath (2001) put forward a different specification for the variable smoothing spline. Because these authors modelled the UK term structure, the tuning parameter definition in steps adopted by Waggoner did not apply. British bond characteristics did not lead to a natural division of the maturity range. These authors then considered a tuning parameter which varies continuously with maturity.

2.2 Nonparametric regression models

Nonparametric regression occupies an intermediate position in the spectrum of modelling techniques, situating itself between the spline models discussed in the previous section and parametric models. The latter typically rely on specific formulae to describe the underlying data generating process. Like spline models, nonparametric regression models hinge upon the critical balance between achieving a good fit (minimising bias) and maintaining smoothness (reducing variance). As James et al. (2021, p.304-306) pointed out, local regression is similar to splines but allows regions used in the estimation to overlap. The choice of the bandwidth, or the span, used for smoothing is pivotal in this regard, as it can lead to models that either connect all observations (overfitting the data) or converge towards linear regression. While the bandwidth may be set based on some prior knowledge about the data, it can also be determined using cross-validation.

This subsection presents two conventional approaches to nonparametric regression: kernel regression and local regression. Although these methods have not seen widespread adoption in the yield curve estimation literature, there is no intrinsic reason not to consider their application. Nonparametric regression techniques are aptly described as *smoothers* because they aim to approximate an unknown function using a smooth curve, aligning closely with the central objective of yield curve estimation.

2.2.1 Kernel regression

In contrast to spline models that partition the domain and estimate a polynomial for each segment, nonparametric regression employs a sliding window over the domain to estimate local trends. At each point (x_0) in the sample, i.e. for each window definition, it assigns different weights for the remainder of observations (x_i, $i \neq 0$). These weights are determined by a weighting function known as the kernel, $K_\lambda(x_0, x_i)$.

The parameter λ defines the bandwidth around x_0 considered in the estimates.[6] It specifies how *local* the estimates will be. Observations within the bandwidth will be weighted according to their distance to x_0, while observations outside the bandwidth will have a weight of zero.

The problem of yield curve estimation can be stated as a nonparametric regression model, such as $Y = g(m) + \epsilon$, where the pairs of interest and maturity (Y_i, M_i) are random variables. Assuming $g(.)$ is a smooth function, which is appropriate for the yield curve, it can be nonparametrically estimated using kernel methods. In this case $g(m)$ would be obtained as the conditional mean of Y given $M = m$:

$$g(m) = E(Y|M = m) = \int y f_{Y|M=m}(y) dy = \frac{\int y f_{M,Y}(m, y) dy}{f_M(m)}. \tag{10}$$

The problem amounts to the task of obtaining estimates for the joint probability density function, $f_{M,Y}(m, y)$, and the marginal probability density function of M, $f_M(m)$, all without making any assumptions about their specific functional forms. Considering the empirical CDF, $\hat{F}_n(m) =$

[6] In addition to defining the bandwidth, a statistically appropriate kernel should exhibit the following characteristics: it must be a nonnegative, bounded function that is normalised (with $\int K(v)dv = 1$), symmetric ($K(v) = K(-v)$), and its use should not reduce a random variable to a constant (with $\int v^2 K(v)dv = \kappa_2 > 0$), see Li and Racine (2007, p. 9).

$\frac{1}{n}\sum_{i=1}^{N}\mathbb{1}_{m_i<m}(m_i)$, these densities can be estimated as

$$\hat{f_M}(m) = \frac{\hat{F}_n(m+h) - \hat{F}_n(m-h)}{2h} = \frac{\sum_{i=1}^{N}\mathbb{1}_{[m-h,m+h]}(m_i)}{2hn} \tag{11}$$

$$= \frac{1}{nh}\sum_{i=1}^{N}\underbrace{\frac{\mathbb{1}_{[m-h,m+h]}(m_i)}{2}}_{\text{(I)}}. \tag{12}$$

The term (I) in equation (12) functions as a kernel (*uniform rectangular window*). It sets the bandwidth centred around the value m with a width of $2h$ and assigns a constant weight of $1/2$ to all observations within this bandwidth. However, an alternative approach involves using a kernel with weights that gradually decrease as you move further from the centre. This approach leads to what is known as the Nadaraya-Watson Kernel Regression, or Local Constant Kernel Estimation, (Li and Racine, 2007, p. 60-64):

$$\hat{g}(m) = \frac{\int y f_{X,Y}(x,y)dy}{f_X(x)} = \frac{\sum_{i=1}^{N} y_i K_\lambda\left(\frac{m_i-m}{\lambda}\right)}{\sum_{i=1}^{N} K_\lambda\left(\frac{m_i-m}{\lambda}\right)}. \tag{13}$$

Several kernel functions can be employed with the Nadaraya-Watson estimator, including the Epanechnikov ($K(v) = \frac{3}{4}(1-v^2)$) , Gaussian ($K(v) = \phi(v) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}v^2}$), and tricube ($K(v) = (1-|v|^3)^3$) kernels, among others. Both the definition of the bandwidth and the selection of the kernel function have a significant impact on the outcomes of these estimates.

2.2.2 Local regression (LOESS)

The Nadaraya-Watson kernel estimator exhibits bias at the boundaries. Moreover, estimates may be biased at interior points when the observations are not uniformly distributed across the domain. The bandwidth is notably asymmetric at the boundaries, leading to bias. At interior points, the bias arises from the disproportionate entry and exit of observations within the bandwidth. One approach to mitigate this bias is to locally fit a linear regression instead of relying on a local average.

The LOcally WEighted Scatterplot Smoothing (Lowess), also referred to as LOcal regrESSion (LOESS), was introduced by Cleveland (1979) as a technique for smoothing scatterplots. This method involves the local estimation of a polynomial regression (linear, quadratic, or cubic) for each data point in the sample. Unlike kernel regression, which uses a fixed distance (λ) to define the bandwidth, LOESS considers a fixed proportion of points to compute local regressions (i.e. the bandwidth may change its size for each observation). The span ($s = k/n$) determines which points will be used in each step, and weights are assigned based on a kernel function. The span serves to balance the trade-off between bias and variance in a manner akin to how the number of knots operates in regression splines or how the smoothing parameter functions in smoothing splines, (Berk, 2016, p. 89).

Considering a linear specification, the LOESS implementation would follow algorithm below.

Local regression algorithm

- For each m_i following $i = 1, \dots, N$

    1. Select $s = (k/n)\%$ of the sample, which corresponds to the m_j points which are nearest to m_i.
    2. Assign a weight $K(m_i, m_j)$ to each of these points such that the nearest point has the higher weight, while the farthest has weight zero. Assign weight zero to points outside the subsample.
    3. Estimate a weighted least squares regression considering the weights assigned in step 2:
    $$\min_{\beta_0,\beta_1} \sum_{j=1}^{N} K(m_i, m_j)(y_j - \beta_0 - \beta_1 x_j)^2$$
    4. Get the estimate: $\hat{m}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

2.3 Nelson-Siegel family models

The polynomial regression discussed previously seeks to model the entire maturity domain using a single function. However, obtaining a satisfactory fit with polynomial regression often requires a highly complex model. This complexity can be computationally expensive to estimate and may entail undesirable properties. In contrast, spline models, while adept at providing a good data fit, entail estimating a series of regressions across the domain. These models yield parameter sets that are not easily summarised or interpreted in economic terms.

This section introduces a family of models, beginning with the specification by Nelson and Siegel (1987), designed to represent the term structure of interest rates with a *parsimonious* function in terms of parameters. These models can effectively depict various yield curve shapes. Furthermore, they allow for a direct association of a small number of parameters with underlying economic features, facilitating the interpretation of their behaviour. These features have contributed to the widespread popularity of these models among central banks worldwide. According to a survey by BIS (2005), most central banks utilise Nelson-Siegel's or Svensson's specifications. Additionally, a more recent account by Nymand-Andersen (2018) notes that the European Central Bank also employs these specifications.

2.3.1 The Nelson-Siegel model

The model proposed by Nelson and Siegel (1987) is rooted in the expectation theory of the term structure of interest rates. This theory motivates the exploration of functions that could serve as solutions to differential and difference equations, generating various yield curve shapes. Thus, these authors introduced a function, presented in Equation (14), which comprises a constant term and a Laguerre function, which is characterised by a polynomial multiplied by an exponential decay term, as follows[7]:

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1-e^{-m/\lambda_t}}{m/\lambda_t}\right) + \beta_{2,t}\left(\frac{1-e^{-m/\lambda_t}}{m/\lambda_t} - e^{-m/\lambda_t}\right) + \epsilon_{m,t}. \tag{14}$$

In Equation (14), three factors ($\beta_{0,t}$, $\beta_{1,t}$, $\beta_{2,t}$) and their factor loadings (the exponential decay terms) contribute to the model's formulation. The factors and loadings in this equation play distinct roles in shaping the yield curve. Specifically:

- Factor $\beta_{0,t}$ is associated with the long-run interest rate and controls the overall level of the function.
- Factor $\beta_{1,t}$ is linked to the short-run interest rate and governs the function's inclination. The factor loading makes it most influential when $m = 0$, decaying monotonically afterwards.
- Factor $\beta_{2,t}$ shapes medium-term interest rates and controls the curvature of the function. The exponential decay parameter (λ_t) determines the maturity at which it exerts the most influence.

The parameter λ_t regulates the decay speed for the factor loadings (for $\beta_{1,t}$ and $\beta_{2,t}$). It's important to note that λ_t is not scale-free, and its value depends on how maturity is measured (e.g., days, months, or years).

According to Nelson and Siegel (1987, p. 475), the theoretical motivation suggests considering different decay parameters for short- and medium-term factors. However, their analysis argues that including a second decay factor in Equation (14) results in over-parametrisation without substantial gains in model fit.

Treating λ_t as an unknown parameter alongside the factors (βs) implies that Equation (14) is nonlinear, necessitating estimation through nonlinear least squares. Nelson and Siegel (1987) recommends exploring a grid of λ_t values and employing ordinary least squares to determine the factors. The optimal λ_t value is then chosen based on the best-fitting estimate among those in the grid.

[7]The specification presented in Equation (14) refers to the yield-to-maturity TSIR representation; it is possible to obtain equivalent specifications for the other forms.

In their original analysis, Nelson and Siegel (1987, p. 480) observed that despite their specification yielding a good fit across various samples, the resulting fitted curve generated non-random residuals displaying dependence on maturity. Consequently, they opted not to investigate the statistical significance of coefficient estimates. Subsequent literature has similarly overlooked this aspect, concentrating primarily on evaluating the model's fit.

Estimating Nelson-Siegel parameters using nonlinear or ordinary least squares methods presents several challenges, as highlighted by Gilli et al. (2010). Two primary issues emerge: the *collinearity problem* and the *optimisation problem*. The former is attributed to the calculation of factor loadings, and the latter is inherent in the nature of the optimisation problem.

Specifically, when applying the factor loadings formulae, numerous combinations of maturities and decay parameter values (λ_t) can lead to pronounced collinearity among the loading factors. While the correlation of factor loadings does not directly influence the fit of the linear regression model, it does impact parameter estimates. As emphasized by Greene (2003, p. 56-57), the effects of multicollinearity include minor fluctuations in the data leading to substantial variations in the estimates, individual coefficients potentially having high standard errors (resulting in low significance) even when all coefficients are jointly significant, and the possibility of coefficient estimates exhibiting an inverted signal concerning the underlying theory, coupled with inaccuracies in their magnitude.

The instability in parameter estimates poses a significant challenge when attempting to align these estimates with economic concepts. For instance, associating the short-run interest with the sum of two coefficients, expressed as $y(0) = \beta_{1,t} + \beta_{2,t}$, becomes problematic in the presence of inaccurate estimates. Furthermore, in the context of the dynamic Nelson-Siegel model proposed by Diebold and Li, it seems improbable for substantial variations to occur across consecutive days in the sample, such as $\beta_{1,t+1}$ differing significantly from $\beta_{1,t}$.

Despite the acknowledged collinearity problem, literature suggests that estimating Equation 14 using OLS, i.e., by fixing $\lambda_t = \lambda$, results in more stable trajectories for the coefficients over time.

To address the collinearity problem, Annaert et al. (2013) introduced a three-step estimation procedure. This approach involves determining the decay factor through a grid search. In cases where the resulting loading factors exhibit high correlation, the authors employ a ridge regression approach to fit Equation 14. By employing this procedure, they observed more stable time paths for the coefficients across the days in their sample.

The *optimisation problem* manifests itself when estimating all parameters through nonlinear least squares. Minimising the sum of squared residuals using Equation 14 proves to be an ill-conditioned problem – it lacks convexity, and the surface of the sum of squared residuals exhibits multiple local minima. Furthermore, the optimisation needs to account for constraints on the parameters in order that the estimates maintain economic meaning. While it is possible to achieve a good fit with NLS, the instability of parameter estimates persists in this scenario as well.

The *optimisation problem* results in many works reporting *numerical difficulties* when implementing NLS for the Nelson-Siegel estimation. The approaches to handle these problems encompass using a global and a local search algorithm (Bolder and Stréliski, 1999), the use of a genetic algorithm for the optimisation (Franklin Jr. et al., 2012), the use of a heuristic for the optimisation (Gilli et al., 2010, 2019), and strategies for defining the initial guesses and constraints (Wahlstrøm et al., 2022).

The following specifications share the same estimation difficulties, enhanced by more parameters.

2.3.2 The Bliss model

According to Bliss (1996, p. 11-12), the over-parametrisation identified by Nelson and Siegel in equation (14) was due to the inclusion of relatively short-maturity bonds (US Treasury bills) in their sample. Bliss addressed this issue by considering bonds with longer maturities, thereby mitigating the over-parametrisation previously observed. Notably, Bliss achieved superior results using a model featuring five parameters – specifically, incorporating two distinct decay parameters – as opposed to the original specification.

While Bliss termed this modified approach the "*extended Nelson-Siegel method*", it has been alternatively referred to as the Bliss model or Bliss specification in other contexts. The formulation is illustrated in equation (15),

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1 - e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}}\right) + \beta_{2,t}\left(\frac{1 - e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}} - e^{-m/\lambda_{2,t}}\right) + \epsilon_{m,t}. \tag{15}$$

Bliss estimated the above equation through nonlinear constrained optimisation. The constraints were stated in relation to the discount function, with the aim of ensuring non-negative forward rates and a positive discount rate in both the short and long ranges.[8]

2.3.3 The Svensson model

Studying a particularly turbulent period on Swedish economy, Svensson (1994) noticed that the Nelson and Siegel (1987) specification was not capable of capturing the yield curve shape in his data.

[8] $d_t(m) \geq d_t(m+1) \Rightarrow e^{-\tilde{y}_t(m)\times m} e^{-\tilde{y}_t(m+1)\times(m+1)} \forall m \leq m_{max}$, and $y(m_{min}) \geq 0$, $y(\infty) \geq 0$.

Therefore, he proposed a specification – equation (16) – with a second medium-term factor, with a specific decay parameter. His intent was to model a second "hump" on the yield curve,

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}}\right) + \beta_{2,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}} - e^{-m/\lambda_{1,t}}\right) + \beta_{3,t}\left(\frac{1-e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}} - e^{-m/\lambda_{2,t}}\right) + \epsilon_{m,t}. \quad (16)$$

2.3.4 The Five-factors model

A further specification in the Nelson-Siegel family considered in this text is the Five-factors model. Two variants of the five-factor models were proposed in different contexts, Björk and Christensen (1999) investigated what parametric representations of the forward rate curves would be consistent with arbitrage-free interest rate models.[9] In particular, they were concerned with Heath et al. (1992) no-arbitrage formulation. Noting that the Nelson-Siegel specification is inconsistent with the arbitrage-free assumption, they developed a variant with five factors which is consistent with the Heath–Jarrow–Morton model.

To increase the flexibility and fit of the Nelson-Siegel family, Rezende (2011) and Rezende and Ferreira (2013) proposed a specification with two short-term components and two medium-term components that decay accordingly to different parameters. While both variants have five factors, they differ slightly in how maturity influences the factor loading on the fifth factor. For the model assessment exercise, we consider only the Rezende and Ferreira specification:

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}}\right) + \beta_{2,t}\left(\frac{1-e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}}\right) + \beta_{3,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}} - e^{-m/\lambda_{1,t}}\right) + \beta_{4,t}\left(\frac{1-e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}} - e^{-m/\lambda_{2,t}}\right) + \epsilon_{m,t}. \quad (17)$$

3 Data and estimation

The ideal source of information for the term structure of interest rates is government bonds' prices (or rates) with various maturities, given their status as the closest option to a risk-free asset with high liquidity. However, time series data for the government bonds' secondary market are available through

[9] An arbitrage opportunity arises when a positive pay-off is guaranteed with no net investment required. An illustrative example is borrowing money at a rate b and simultaneously lending it at a rate c, where $c > b$, all without incurring operational costs, (Dybvig and Ross, 1989). An arbitrage-free interest rate model imposes some constraints in the mathematical representation of the term structure in order that the behaviour of the instantaneous forward rate curves offers no arbitrage opportunities.

proprietary databases. For example, Anbima provides free data on the transactions of the last five working days. [10] For more extended periods, the data is exclusively available through subscription.

As a result, most empirical studies on the Brazilian term structure of interest rates turn to data from One-day Interbank Deposit Futures ("DI1")[11] contracts provided by B3[12]. These data represent the daily average of one-day interest rate futures contracts, expressed as a percentage rate per annum compounded daily based on a 252-day year. Due to the specific settlement characteristics of these futures contracts, they are often considered close to being risk-free and exhibit significant liquidity. Notably, the interest rates specified in DI1 contracts correspond to spot rates (Berger, 2015, p. 71-72), making them suitable for estimating the term structure of interest rates in its yield-to-maturity form.

Each working day, B3 releases two to five versions of its *Price Report*, documenting all transactions during that day. Historical data from the Price Report is available in XML files, which were downloaded from 2 January 2018 (the oldest report available) to 22 April 2023, using the R package `RSelenium`. This timeframe encompasses 1313 working days, as calculated by the R package `bizdays`, using the calendar "*Brazil/ANBIMA*."

For each day, we considered the last Price Report release, except for 4 April 2021, where the last released XML file was corrupted. For this day, we utilised the second-last file. The XML files were parsed using R's `XML` package, extracting information solely about DI1 transactions. Typically, a day has 37 or 38 records, each corresponding to a different DI1 maturity. Even though there are some variables related to DI1 transactions each day we used two key pieces of information: the date when the contract is due ("*data de referência*") as maturity, and the adjusted rate ("*taxa ajuste*") as the interest rate level. During the analysis it was detected that data from 2 December 2022 had its first maturity equal to "`-1`" and the correspondent yield equal to "`NA`". We excluded only this pair maturity-yield from this day data.

3.1 Estimation procedures

All estimates and plots were performed using the R software. The functions, algorithms, and their options used in the estimation of the different models are documented below:

- **Polynomial Regression**: this model was implemented using the `lm()` and `poly()` functions. A rule of thumb was adopted to determine the polynomial degree, employing degree 4 for a

[10] Ambima - taxas de títulos públicos.

[11] Among others, that is the case of Fraletti (2004), Caldeira (2011), Franklin Jr. et al. (2012), and Caldeira et al. (2016).

[12] B3 - Histórico / Boletins Diarios / Pesquisa por Pregão.

normal curve and degree 9 for inverted or humped curves.

- **McCulloch Cubic Splines**: the knots were defined according "*McCulloch's rule*" using the `quantile` function on each day's maturity vector. Then the natural cubic splines basis was generated using function `ns()` from the `splines` package. Finally the splines were estimated using the `lm()` function.

- **Smoothing Splines**: this model was implemented using the `ss()` function from the `npreg` package, (Helwig, 2022). The smoothing parameter was defined by generalised cross-validation.

- **Kernel Regression**: this model was implemented using the `npreg()` function from the `np` package, using the local constant option to obtain the Nadaraya-Watson estimator. We considered one specification with a fixed bandwidth (0.3) and other with the bandwidth defined by cross-validation, for the latter we allowed the process of optimising the cross-validation function to restart twice from different (random) initial points.

- **Local regression - LOESS**: this model was implemented using the `LOESS()` function from the `stats` package with a span fixed in 0.2.

- **Nelson Siegel model**: this model was implemented in two ways, firstly fixing $\lambda_t = 368$ following Diebold and Li and estimating the model by OLS. Secondly, the model was implemented using nonlinear least squares with the function `nloptr` from the package with the same name (Ypma et al., 2022), utilising the "Improved Stochastic Ranking Evolution Strategy" algorithm for optimisation and a convergence criterion of 10^{-6} limited to 15,000 iterations. This algorithm was chosen because it allows inequality restrictions on the parameter values, two restrictions were imposed $\beta_{0,t} > 0$ and $\beta_{0,t} + \beta_{1,t} > 0$. The initial guesses were made according Wahlstrøm et al. (2022) recommendation: $\beta_{0,t}$ equal the average yield from the three shorter maturities, $\beta_{1,t}$ equal first maturity yield minus the average yield from the three shorter maturities , $\beta_{3,t}$ equal zero, and λ_t equal 368.

- **Bliss, Svensson, and the Five-factor models**: these models were implemented using the nonlinear Nelson-Siegel approach, employing the `nloptr` function with the "Improved Stochastic Ranking Evolution Strategy" algorithm. The primary distinction lies in the expansion of the initial guesses vector. For the Svensson specification an additional restriction was included in the optimisation: $\lambda_{2,t} > \lambda_{1,t}$. For the Five-factor model, we allowed 30,000 iterations, while the others were limited to 15,000 iterations.

4 Methodology

To compare estimates from the models presented in the previous section, it is indispensable to define metrics that enable the evaluation of various dimensions by which a yield curve estimate can be assessed. Relevant literature[13] consistently considers dimensions such as flexibility (goodness-of-fit), robustness (stability to outliers), and smoothness.

When evaluating **goodness-of-fit**, the aim is to measure how well the underlying method fits each data point in the sample. This assessment can be conducted using metrics such as Mean Squared Error (MSE) or Mean Absolute Error (MAE):

$$MSE = \frac{\sum_{i=m_1}^{M} (y_i - \hat{y}_i)^2}{M} \qquad\qquad MAE = \frac{\sum_{i=m_1}^{M} |y_i - \hat{y}_i|}{M}. \tag{18}$$

However, as most estimation procedures presented hinge on minimising the sum of square errors (SSE), a straightforward comparison of metrics like the MSE would inherently favour more parametrised models over their more parsimonious counterparts. In the context of yield curve estimation, relying solely on this metric in the **full sample** would not accurately portray the actual performance of the models, (Bliss, 1996). Moreover, the yield curve's short end is typically more populated (with more bonds or derivative contracts) than the long-range, and usually, it has a more complex form. Thus, considering a single fit metric for the whole maturity range could be misleading. A model with an average performance over all maturities could be levelled with others which systematically perform poorly at a specific range.

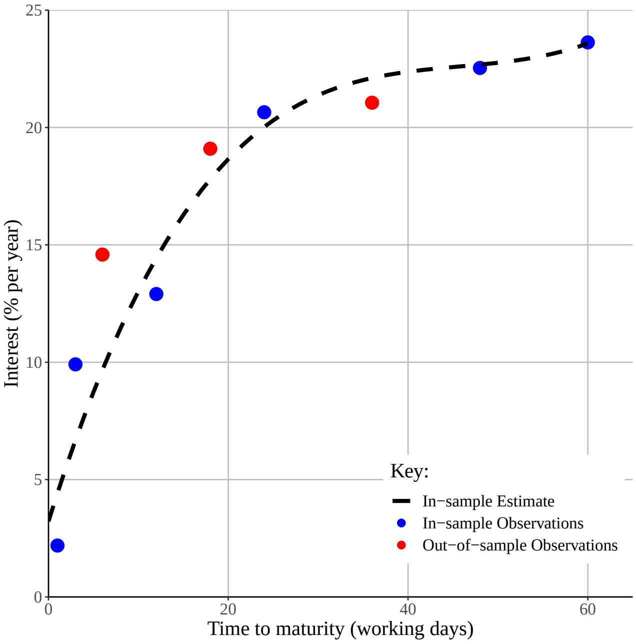
To overcome these difficulties, we consider a **training set** (or in-sample set) and a **validation set** (or out-of-sample set) for each day in the sample. A typical working day has around 38 data points; we classify these points into three groups according to maturity: maturities up to one year, maturities above one year and below three years, and maturities over three years. For each of these groups, we randomly select two data points, ensuring that the validation set comprises six observations (roughly 15% of all observations in that day), consisting of two observations from each group. The only restriction in this process was that the shortest and the longest maturities on a given day should be in the training set.

All models are estimated for each day in the sample, utilising their respective training sets. Subsequently, the performance of in-sample estimates is evaluated by assessing how well they fit in- and out-of-sample data using MSE and MAE. This evaluation considers both the entire maturity range

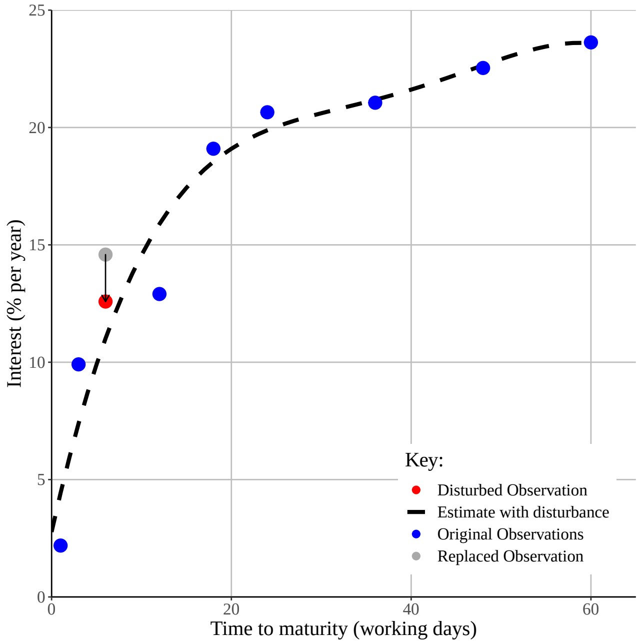
[13] Noteworthy works, including Bliss (1996), Seppälä and Viertiö (1996), Anderson and Sleath (2001), and Nymand-Andersen (2018), explore these dimensions, albeit with some differences in the specific measures employed.

and its divisions. Figure 1a illustrates this process with synthetic data: the blue dots represent the in-sample data used to estimate the dashed line, while the red dots represent the out-of-sample data. In-sample performance is gauged by comparing the distance between the blue dots and the dashed line, whereas out-of-sample performance is evaluated by comparing the distance between the red dots and the dashed line.

Figure 1: An illustration of goodness-of-fit and robustness assessment procedures.



(a) Out-of-sample points



(b) Estimation after a disturbance

Note: Synthetic data used for illustration. **Source**: Elaborated by the author.

The **robustness** evaluation aims to assess how the estimates respond to a disturbance in the interest rates of a specific maturity. In essence, it seeks to understand how the estimates behave in the presence of outliers or measurement errors. To conduct the robustness evaluation, a hypothetical data set is constructed for each day. In this set, a randomly selected maturity (excluding the first and the last) has its interest rate disturbed by an increase or decrease of 2%. Subsequently, each model is estimated using the data with the disturbance, and the estimated curve is used to calculate the MSE and MAE, considering the original data while excluding the disturbed point. An ill-conditioned[14] model would exhibit a significant increase in these measures, indicating that the disturbance has affected the estimates across multiple maturities. Figure 1b illustrates this process with synthetic data. The third

[14]"*A mathematical problem is called well-conditioned provided that small changes in the data leads only to small changes in the (exact) solution. If this is not the case, we call. the problem ill- conditioned.*", Hämmerlin and Hoffman (1991, p. 20-21)

maturity was selected in this case, and its interest rate level was reduced by 2%. Subsequently, the model was estimated considering both the red and blue dots. The resulting dashed line represents the estimated curve, and goodness-of-fit measurements were calculated using the grey and blue dots.

Smoothness is a valued quality for two crucial reasons. Firstly, many applications relying on yield curve estimates depend on their differentiability; consequently, an estimate with kinks at numerous maturities would be of limited utility. Secondly, an excessively rough estimate may suggest overfitting of the data, potentially impairing its interpolation ability. Ramsay and Silverman (2005, p. 84) point out that a way to quantify the notion "*roughness*" of a function is to consider the integrated squared second derivative of that function

$$R(y_t) = \int_{m_1}^{M} [D^2 y(s)]^2 \, ds. \tag{19}$$

The roughness measure in equation (19) was adopted by Adams and Deventer (1994, p. 54) and Varga (2009, p. 381) in the context of comparing smoothness of the yield curve estimates.[15] Seppälä and Viertiö (1996, p. 21) argue that since we have more information at the short end than at the long end and expect a more complex behaviour at the short end (McCulloch, 1971), the fact that the roughness measure in equation (19) equally penalises changes in yield curve's slope in any maturity is not appropriate. Thus, these authors propose a modified roughness measure which weights the slope changes by maturity

$$R_2(y_t) = \int_{m_1}^{M} [s \times D^2 y(s)]^2 \, ds. \tag{20}$$

Given the increasing weight of second derivatives at longer maturities, $R_2(.)$ also helps to detect changes in the yield curve steepness after the last observed maturity. In other words, it helps evaluate whether the models' **extrapolation ability** aligns with the underlying economic reasoning.

Alternatively, Waggoner (1997) considers a different modification considering the average roughness instead of the total roughness. When comparing different models this modification makes sense once one is considering different maturity ranges for different models, that is, roughness will not be inflated by a longer maturity range,

$$R_3(y_t) = \frac{1}{M - m_1} \int_{m_1}^{M} [D^2 y(s)]^2 \, ds. \tag{21}$$

We employ these three roughness metrics by computing them daily for each model across a maturity span of 0 to 3800 working days (approximately 15 years). This range, reflective of the average

[15] Nymand-Andersen (2018) adopted a different approach to evaluate smoothness, this author used the spread between the n-period rate and the m-period ($s_t(n, m) = y_t(n) - y_t(m)$) as an indicator. This approach is not used in this text.

most extended maturity observed daily, facilitates the evaluation of the models' extrapolation characteristics.

4.1 Two formal tests

Although the metrics presented above offer ways to compare estimates from different models, determining the best model in one of these dimensions from the repeated daily measures requires clarification. Descriptive statistics, such as the out-of-sample MAE averaged over the days, provide insight into each model's performance. However, relying solely on these measures can be tricky since nothing can be said about the **statistical significance** of the differences. To overcome this challenge, we adopt the approach introduced by Koning et al. (2005) to compare the forecast performance of several models applied to many time series. Although our focus is on interpolation accuracy instead of forecasting accuracy, we face a similar problem of comparing different models' performance.

On each day in the sample, model performance can be ranked according some of the metrics (MAE or smoothness' R, for instance). The Friedman test is a nonparametric test which provides a way to evaluate whether different rankings are equal (H_0), i.e. they correspond to an ordering of i.i.d. random variables for each day, against the alternative hypothesis that the rankings are indeed different (H_A), i.e. each day the ranking is an ordering of independent random variables that indeed differ in location. To compare K models evaluated (ranked) over D days, under H_0 we have the test statistic S:

$$S = \frac{12D}{K(K+1)} \sum_{k=1}^{K} \left(\bar{R}_k - \frac{K+1}{2} \right)^2 \underset{D \to \infty}{\sim} \chi^2_{K-1}. \tag{22}$$

Where $\bar{R}_k = \frac{\sum_{d=1}^{D} R_{k,d}}{D}$ is the average rank position of model k on all days in the sample, and $R_{k,d}$ is the rank position of model k on day d.

The Friedman test is an overall test, and the rejection of H_0 in the Friedman test indicates that one of the rankings is different from the others. However, it does not provide a means to directly compare the models' performance, i.e., to evaluate whether one model is systematically superior to another. Hollander et al. (2014, p. 316-321) present a generalisation of the Friedman test, the Wilcoxon–Nemenyi–McDonald–Thompson test, which allows **multiple pairwise comparisons**.

Let τ_k be the effect of model k on the underlying random variable that determines its position in the ranking on a given day. The WNMT test considers a series of null hypotheses $H_{0,k_1,k_2} : \tau_{k_1} = \tau_{k_2}$, where $k_1 = 1, 2, \ldots, K$ and $k_2 = 1, 2, \ldots, K | k_2 \neq k_1$, against the corresponding alternative hypotheses $H_{A,k_1,k_2} : \tau_{k_1} \neq \tau_{k_2}$ where $k_1 \neq k_2$. The pairs of H_0 and H_A correspond to the possible combinations among the models under analysis. Then, each H_0 is rejected if and only if

$$|\bar{R}_{k_1} - \bar{R}_{k_2}| \geq r_{\alpha,K,D}, \tag{23}$$

where the critical value, $r_{\alpha,K,D}$, is set to make the experimentwise error rate equal to α (Hollander et al., 2014, p. 316; Koning et al., 2005, p. 399). Thus, the $r_{\alpha,K,D}$ is the largest constant such that $P_{H_0}\left((max\bar{R}_k) - (min\bar{R}_k) \geq r_{\alpha,K,D}\right) \leq \alpha$, and the large-sample approximation (large D in the present case) for equation (23) gives

$$r_{\alpha,K,D} \approx q_{\alpha,K}\sqrt{\frac{K(K+1)}{12D}},$$

where $q_{\alpha,K}$ is the upper α percentile of the range of K independent $N(0,1)$ variables.[16]

To implement the multiple comparison tests, we follow Koning et al. (2005) once again, who used plots to present the results compactly. In next section, for each model k, an interval is drawn with length $r_{\alpha,K,D}$ and centred at $\bar{R}_k$. If the intervals for two models do not overlap, H_0 – indicating that both models perform equally in ranking terms – is rejected.

A line is drawn at the upper boundary of the interval of the best model (i.e. that with the lowest average ranking for a given metric). That will be the lowest upper boundary among the models analysed and this reference line corresponds to "*the unconstrained multiple-comparison procedure with the best, deducted from all pairwise comparisons*", (Hsu, 1996, as cited in Koning et al.,2005, p. 400). Therefore, all models with confidence interval above the reference line perform significantly worse than the best model.

To our knowledge, multiple comparison tests have not been used before in evaluating yield curve models. The closest related work is Varga (2009, p. 385-387), which applied the Friedman test pairwise to evaluate yield curve models regarding absolute error performance.

5 Empirical results and discussion

We estimated 11 different specifications for each day over the sample period from 2 January 2018 to 22 April 2023. Each specification was estimated at least three times daily: in-sample estimate, full-sample estimate, and disturbed sample estimate. Here, we compare the models' performance according to the metrics explained before and test whether the performances of the models are statistically different.

[16] $q_{\alpha,K}$ can be obtained in R using the function `cRangeNor()` from the package `NSM3`.

5.1 Goodness-of-fit analysis

In this subsection, we analyse the goodness-of-fit of all models. Starting with descriptive statistics, Table 1 presents the average MAE for the days in the sample, illustrating the in-sample and out-of-sample fit and how each model adapts to different maturity ranges. Similarly, Table 2 shows the average MSE.

From these averages, nonparametric and local regression models systematically fit better than the parametric specifications. On average, the smoothing spline has the best in-sample fit. However, considering out-of-sample performance, the LOESS outperforms even the smoothing spline on all different ranges considered.

Both Kernel regression specifications fit the out-of-sample data poorly, especially at the short range. Considering the Nelson-Siegel family, the baseline Nelson-Siegel specification estimated by OLS has the worst performance when estimated with all maturities and on the short-range sub-sample. The other variants have very similar performance according to the descriptive statistics.

When assessing long-range performance, several models demonstrate a comparable fit. McCulloch's Splines, Smoothing Splines, and LOESS exhibit similar performance in terms of MAE or MSE. Similarly, within the Nelson-Siegel family, the models exhibit comparable results.

Even though the descriptive statistics provide some insight into the models' relative performance, it is challenging to establish whether one systematically outperforms the others. For this kind of evaluation, firstly, we perform a Friedman rank test, considering the daily MAE rankings. Results shown in Table 3 confirm that the models have different performances. Then, we implement the **multiple comparison procedure** on Figures 2 and 3.

From Figure 2a, it is possible to conclude that the smoothing splines model has the best in-sample performance considering the whole maturity range. However, according to Figure 2b, the smoothing splines and the LOESS are not statistically different in fitting out-of-sample data. Likewise, considering the fit at the short range (Figures 2c and 2d), both models are not statistically different but fit the data systematically better than the alternatives.

Again, for the Medium Range, the smoothing splines model has the best in-sample performance, but the smoothing splines and the LOESS are not statistically different in fitting out-of-sample data (Figures 3a and 3b). Finally, for the Long Range data, the smoothing splines model has the best in-sample performance once more. However, considering out-of-sample fit, the McCulloch Cubic Spline is not statistically different from the smoothing splines, and the latter is not statistically different from the LOESS, (Figures 3c and 3d).

Table 1: Assessing Model Goodness-of-fit: Mean Absolute Error (MAE) in- and out-of-sample

Model	All maturities		Short Range		Medium Range		Long Range	
	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample
Polynomial Regression	0.118840	0.135760	0.160462	0.178433	0.105082	0.106294	0.095933	0.124005
McCulloch CS	0.023644	0.031496	0.032464	0.042469	0.026535	0.032405	0.016373	0.020813
Smoothing Spline	0.003213	0.020296	0.006668	0.024313	0.002209	0.017817	0.001239	0.018908
Kernel Reg. (fixed bdw)	0.095110	0.120710	0.183239	0.205928	0.125325	0.131440	0.021591	0.040902
Kernel Reg. (cv bdw)	0.076550	0.105846	0.143837	0.170425	0.092280	0.111770	0.023380	0.045974
LOESS	0.007157	0.018793	0.008263	0.021981	0.006291	0.017151	0.006733	0.017301
Nelson-Siegel (NLS)	0.111976	0.121711	0.141323	0.157882	0.120244	0.120598	0.089324	0.091536
Nelson-Siegel (OLS)	0.153710	0.162403	0.209852	0.231251	0.138864	0.136730	0.121753	0.124293
Bliss	0.109594	0.121411	0.140560	0.160035	0.120637	0.122806	0.084251	0.087442
Svensson	0.109674	0.117226	0.128766	0.142198	0.114051	0.116925	0.095133	0.096409
Five-factors	0.094492	0.104155	0.116812	0.133174	0.099230	0.105186	0.077559	0.079559

Table 2: Assessing Model Goodness-of-fit: Mean Square Error (MSE) in- and out-of-sample

Modelo	All maturities		Short Range		Medium Range		Long Range	
	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample
Polynomial Regression	0.033964	0.043722	0.063683	0.077969	0.022139	0.021419	0.018339	0.034213
McCulloch CS	0.001443	0.002783	0.002843	0.005690	0.001418	0.002143	0.000472	0.000743
Smoothing Spline	0.000130	0.001318	0.000361	0.002670	0.000026	0.000621	0.000014	0.000747
Kernel Reg. (fixed bdw)	0.036470	0.047228	0.090432	0.113293	0.031550	0.033067	0.001042	0.003195
Kernel Reg. (cv bdw)	0.029296	0.041979	0.072391	0.100171	0.022220	0.028784	0.002114	0.004864
LOESS	0.000160	0.001204	0.000293	0.002521	0.000082	0.000565	0.000099	0.000601
Nelson-Siegel (NLS)	0.024120	0.029407	0.037790	0.048947	0.025528	0.026085	0.014524	0.015234
Nelson-Siegel (OLS)	0.044703	0.051686	0.081522	0.101900	0.033704	0.032094	0.024089	0.025138
Bliss	0.023734	0.029105	0.039264	0.051148	0.025561	0.026139	0.012483	0.013032
Svensson	0.022188	0.025938	0.030879	0.040562	0.022580	0.023584	0.016241	0.015592
Five-factors	0.019317	0.023381	0.030109	0.040414	0.018670	0.020718	0.012301	0.011624

Table 3: Friedman rank sum test - Models ranked by MAE

Specification	Friedman chi-squared S	df	p-value
Full Sample	9382.04	10	< 2.2e-16
In-sample	8777.95	10	< 2.2e-16
Out-of-sample	7791.03	10	< 2.2e-16
In-sample (Short range)	8005.15	10	< 2.2e-16
Out-of-sample (Short range)	6243.95	10	< 2.2e-16
In-sample (Medium range)	8023.11	10	< 2.2e-16
Out-of-sample (Medium range)	6075.48	10	< 2.2e-16
In-sample (Long range)	10330.40	10	< 2.2e-16
Out-of-sample (Long range)	6829.72	10	< 2.2e-16
Sample with perturbed data	5306.14	10	< 2.2e-16

The key takeaway from the multiple comparison procedures is that smoothing splines and the LOESS are not statistically different for out-of-sample interpolation.

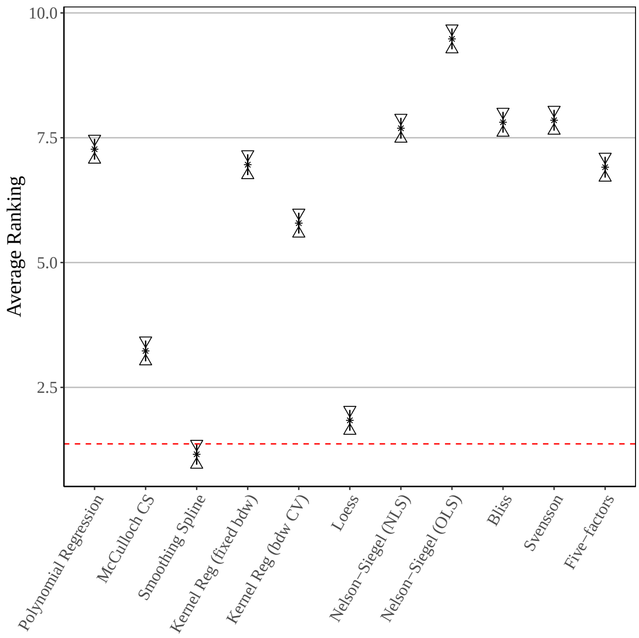
One further analysis can be made about the goodness-of-fit by evaluating how the metrics evolve across the sample. Figures 4 and 5 plot the daily MAE for all models under analysis. All models experience a performance deterioration between mid-2021 and mid-2022. This decline is attributed to the change in the shape of the Yield Curve. Figure 6 provides a glimpse of the Yield Curve dynamics, indicating that the period of increased MAE for all models corresponds to **humped yield curves** – indicating a term structure where the medium range has the highest interest levels.

5.2 Robustness analysis

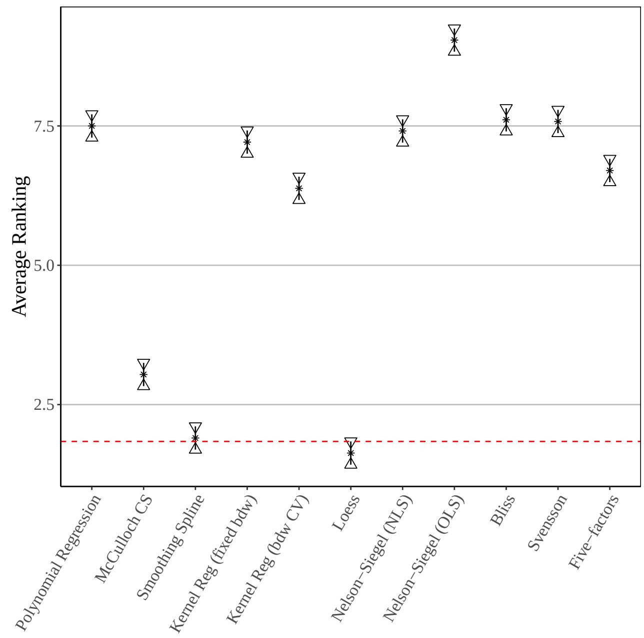
This section analyses how the model fit changes when one outlier is introduced to the sample. Table 4 presents the MAE and MSE average values considering the models estimated using the original and modified data (with a disturbance in a random maturity each day).

The table shows that all models experience a deterioration in their fit, as measured by MAE and MSE. However, the Smoothing Splines model and the LOESS still have the best average fit, even in the presence of an outlier. The Kernel regression specifications show the largest worsening in the fit metrics (in absolute terms). On the other hand, the Nelson-Siegel family models are remarkably robust to outliers.

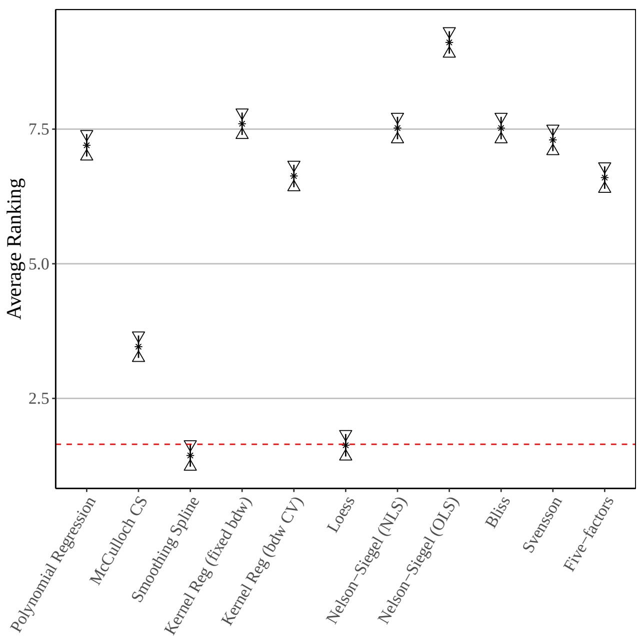
Figure 2: Multiple Comparison Procedure for Goodness-of-Fit - Models Ranked by MAE (Overall and Short Range)



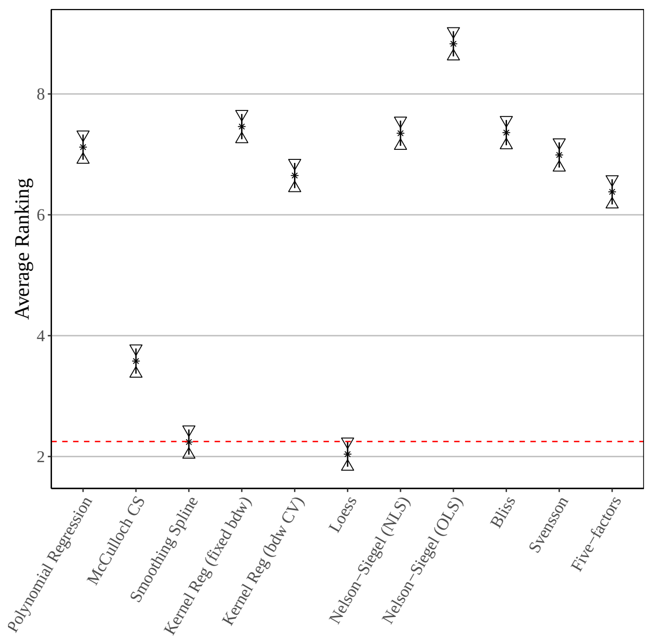
(a) In-sample



(b) Out-of-sample



(c) In-sample Short Range

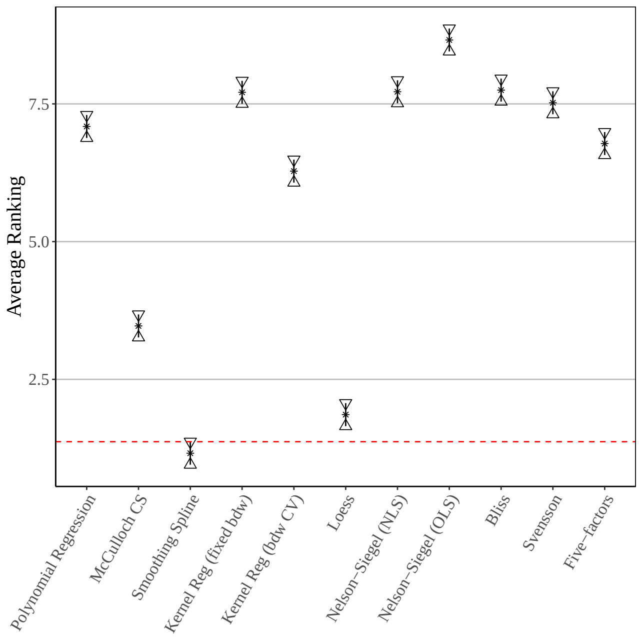


(d) Out-of-Sample Short Range

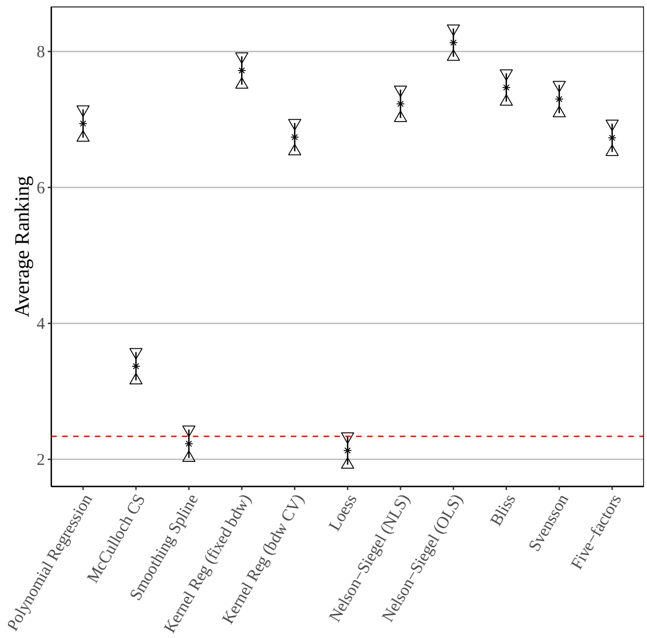
Source: Elaborated by the author.

The Friedman test statistic in Table 3 indicates significant differences among the models in both the original sample and the sample with disturbance. The multiple comparison test in Figure 7 reveals that, except for the Kernel regression specifications, the average rankings of the models remain

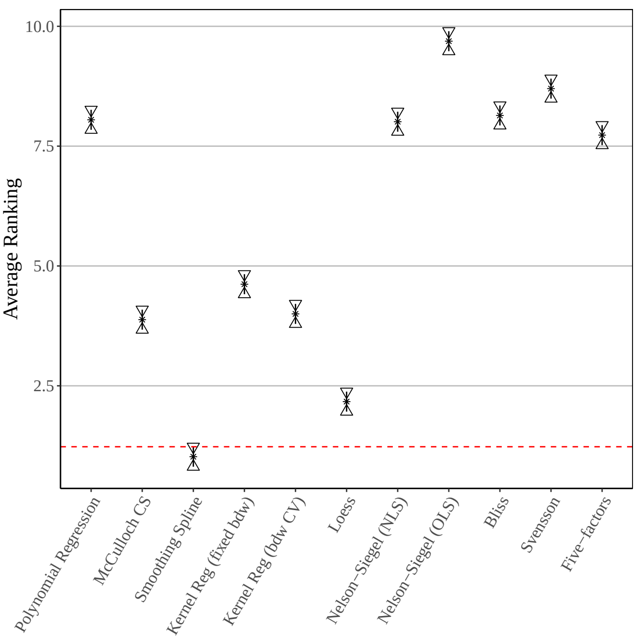
Figure 3: Multiple Comparison Procedure for Goodness-of-Fit - Models Ranked by MAE (Medium Range and Long Range)



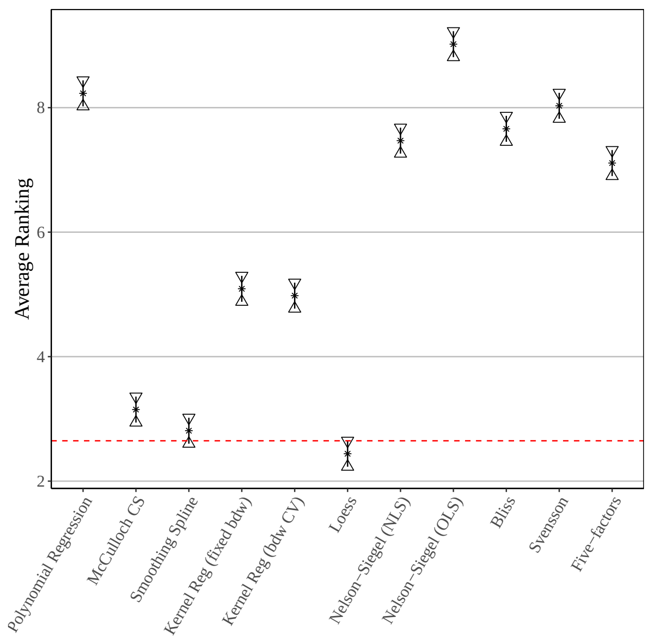
(a) In-sample Medium Range



(b) Out-of-sample Medium Range



(c) In-sample Long Range

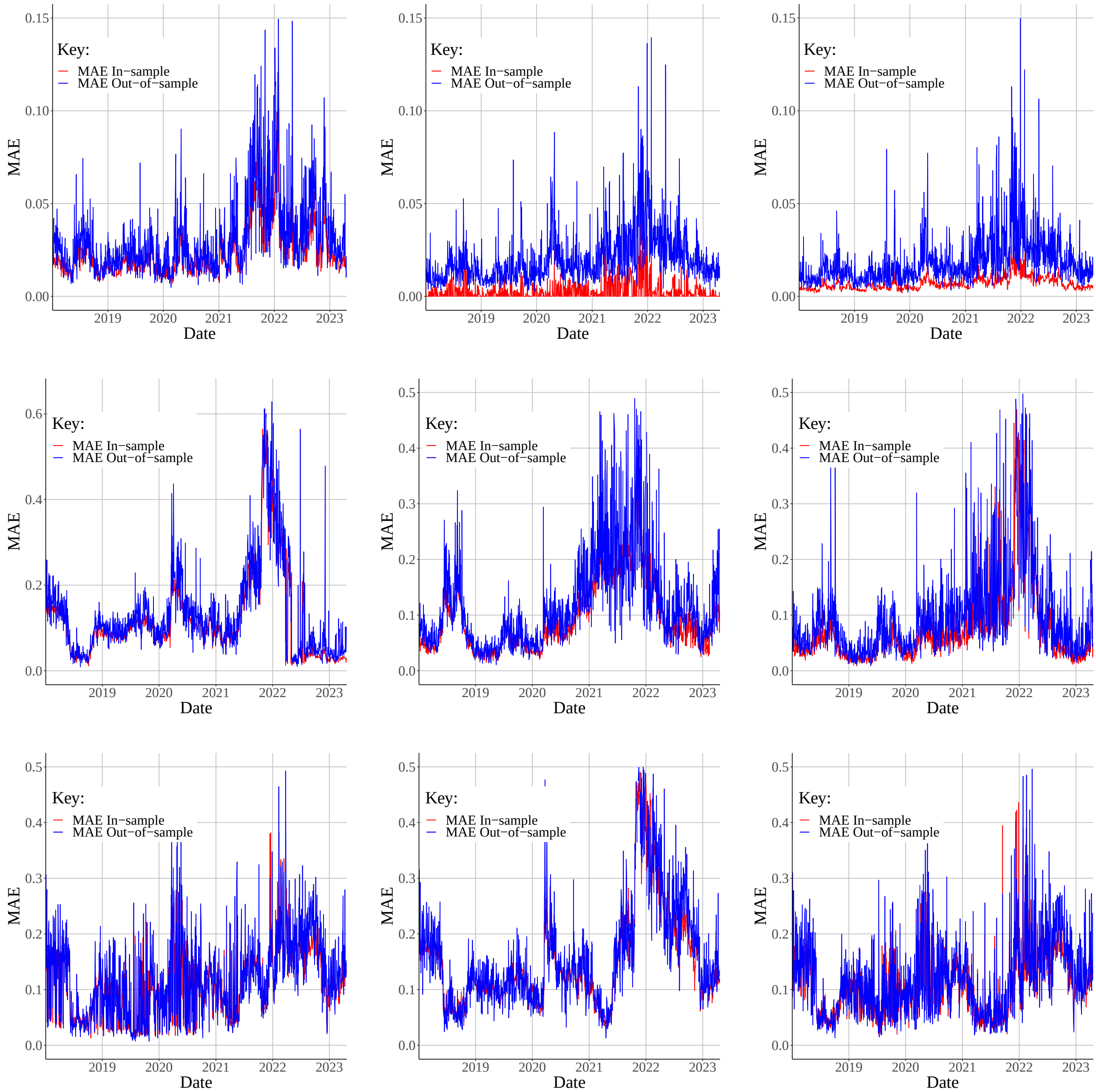


(d) Out-of-Sample Long Range

Source: Elaborated by the author.

unchanged with the introduction of the outlier, and the smoothing spline maintains the best-fit rank in both settings.

Figure 4: In-sample and Out-of-sample Accuracy Across Sample Days (1 of 2)



a b c
d e f
g h i

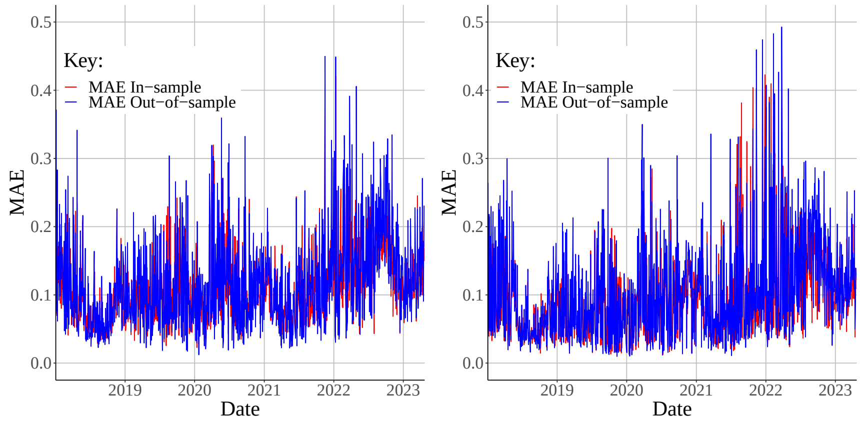
Panels: (**a**) McCulloch Natural Cubic Spline. (**b**) Smoothing Splines. (**c**) LOESS. (**d**) Polynomial Regression. (**e**) Kernel Regression (Fixed bandwidth). (**f**) Kernel Regression (Bandwidth by CV). (**g**) Nelson-Siegel (NLS). (**h**) Nelson-Siegel (OLS). (**i**) Bliss.

5.3 Smoothness analysis

This section analyses how smooth the models' estimates are according to the three roughness metrics presented previously. Table 5 provides the average metrics considering all days in the sample.

Considering R, Polynomial, Svensson, and LOESS specifications present the lower roughness.

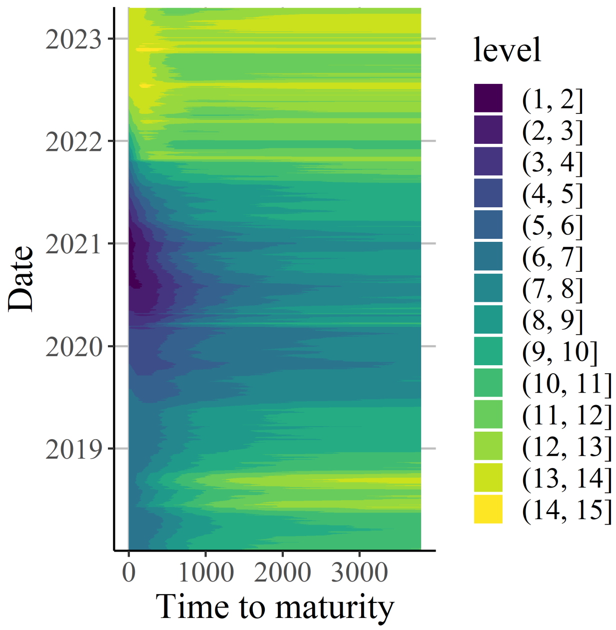
Figure 5: In-sample and Out-of-sample Accuracy Across Sample Days (2 of 2)



(a) Svensson **(b)** Smoothing Splines

Source: Elaborated by the author.

Figure 6: The Yield Curve Evolution: January 2018 - April 2023



Source: Elaborated by the author.

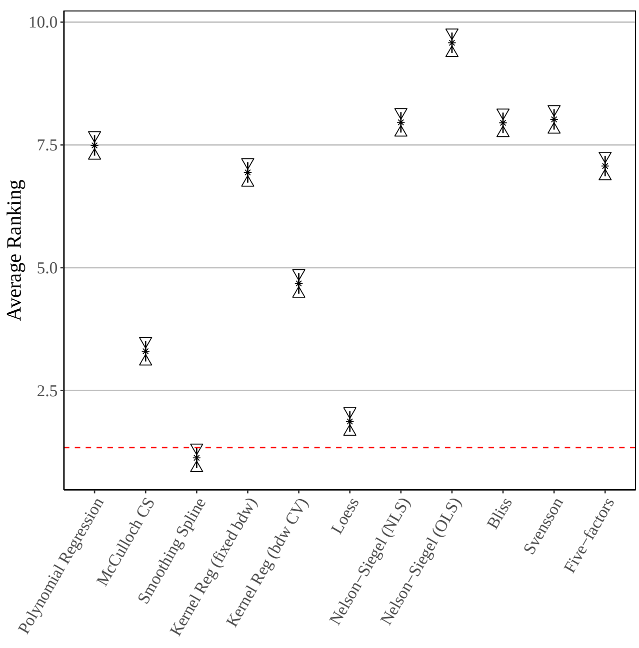
Analysing the average R_3 indicates that the Polynomial and the Kernel regression with fixed bandwidth are the smoother estimates. However, all models present low values for R and R_3, making it hard to tell differences from the averages.

Regarding R_2, a different pattern emerges; the Nelson-Siegel family correspond to the smoother estimates. Conversely, the cubic spline's performance is inferior. This discrepancy arises due to the metric's weighting that penalises steepness at higher maturities. A high R_2 suggests that the yield curve estimate curvature changes substantially for the smoothing splines after the last observed

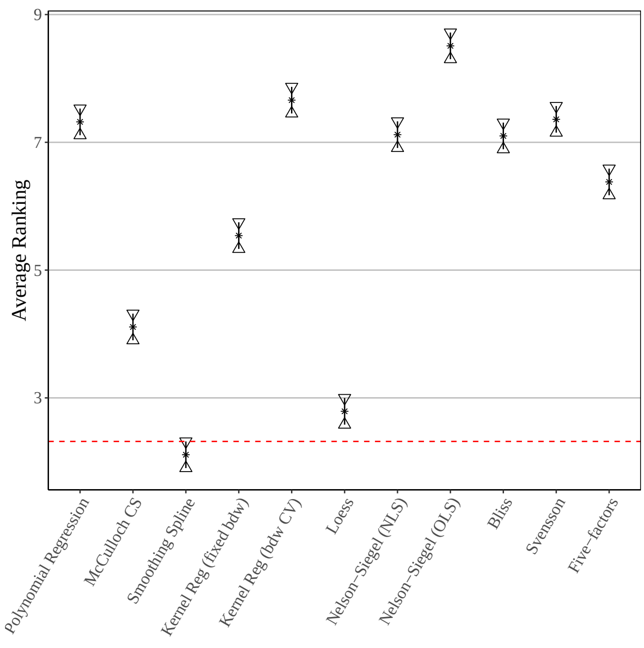
Table 4: Assessing Model Robustness: MAE and MSE in Original and Perturbed Sets

Modelo	MAE		MSE	
	Original data	**Perturbed data**	**Original data**	**Perturbed data**
Polynomial Regression	0.120535	0.145309	0.034326	0.041096
McCulloch CS	0.022069	0.087315	0.001165	0.018092
Smoothing Spline	0.003673	0.067968	0.000146	0.079302
Kernel Regression (fixed bdw)	0.085938	0.113314	0.031320	0.043631
Kernel Regression (cv bdw)	0.046772	0.148206	0.015330	0.066718
LOESS	0.008231	0.076976	0.000204	0.052546
Nelson-Siegel (NLS)	0.114567	0.132499	0.024878	0.029479
Nelson-Siegel (OLS)	0.154595	0.161646	0.044970	0.044966
Bliss	0.109715	0.130170	0.023354	0.029164
Svensson	0.110072	0.132498	0.022748	0.029438
Five-factors	0.093842	0.121020	0.018988	0.026120

Figure 7: Multiple Comparison Procedure for Robustness - Models Ranked by MAE



(a) Original Sample



(b) Sample with disturbance

Source: Elaborated by the author.

maturity. In other words, a high R_2 for the smoothing spline indicates poor extrapolation after the last observation.

Table 5: Assessing Model Smoothness: Three Roughness Measures

Model	R	R_2	R_3
Polynomial Regression	8.563060E-08	6.874353E-01	2.253437E-11
McCulloch CS	3.882038E-07	1.925056E-02	1.021589E-10
Smoothing Spline	3.093370E-03	3.942163E+04	8.140449E-07
Kernel Regression (fixed bdw)	1.347686E-07	2.775508E-01	3.546542E-11
Kernel Regression (cv bdw)	2.406275E-06	1.534375E+01	6.332302E-10
LOESS	7.055688E-06	5.557780E-01	1.856760E-09
Nelson-Siegel (NLS)	1.408724E-06	1.241196E-02	3.707169E-10
Nelson-Siegel (OLS)	1.407544E-06	1.244910E-02	3.704063E-10
Bliss	1.536130E-06	1.224638E-02	4.042447E-10
Svensson	7.260830E-07	1.164079E-02	1.910745E-10
Five-factors	5.578462E-07	1.195935E-02	1.468016E-10

Ranking the models according the smoothness metrics and applying the Friedman tests indicates that models' performance in terms of roughness are indeed different.[17]

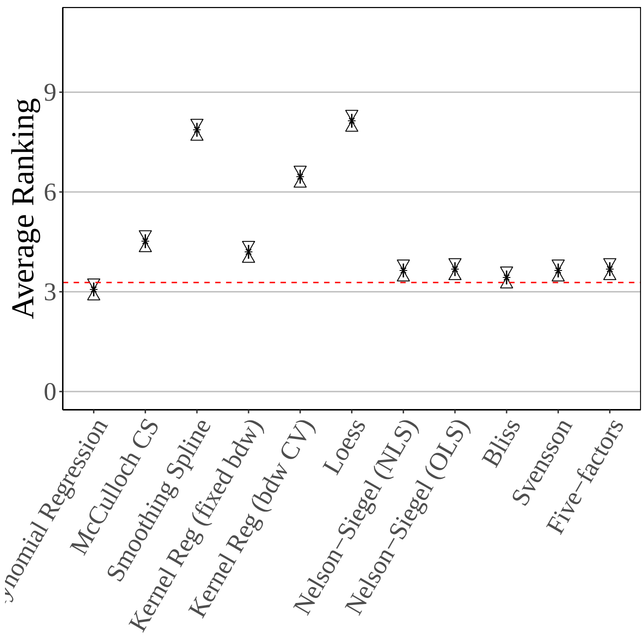
Figure 8 shows the **multiple comparison procedure**. Estimates within the Nelson-Siegel family are not statistically different in any setting. Considering the rank by R, the Nelson-Siegel family and the Polynomial regression systematically have the lowest roughness in the sample. Considering R_2, the Nelson-Siegel family is unchallenged, while the rankings given by R_3 show the Polynomial regression with the fewest rough estimates.

6 Conclusion

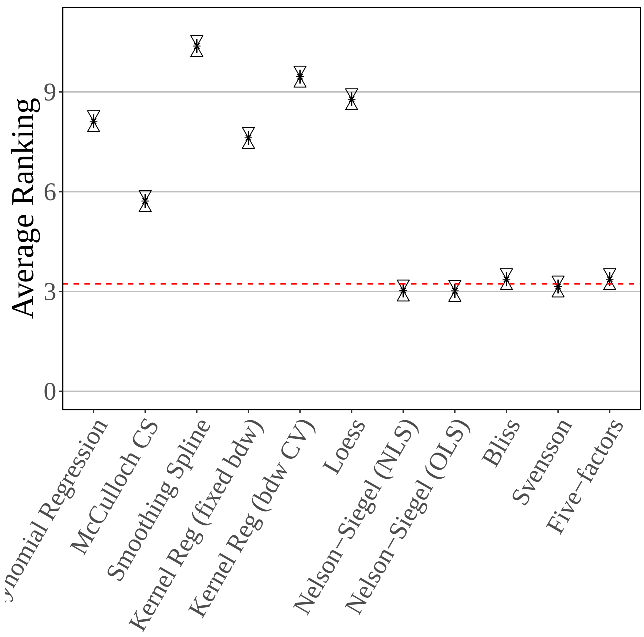
This text provides a comprehensive evaluation of empirical models of the Term Structure of Interest Rates. Literature usually considers nonparametric or spline models in addition to the parsimonious function models, derived from Nelson and Siegel (1987)'s seminal work.

[17]The Friedman statistics are: 5915.6, when the models are ranked by R; 10228, when the models are ranked by R_2; and 10823, when the models are ranked by R_3. In all cases with 10 degrees of freedom and a p-value smaller than $2.2e^{-16}$.

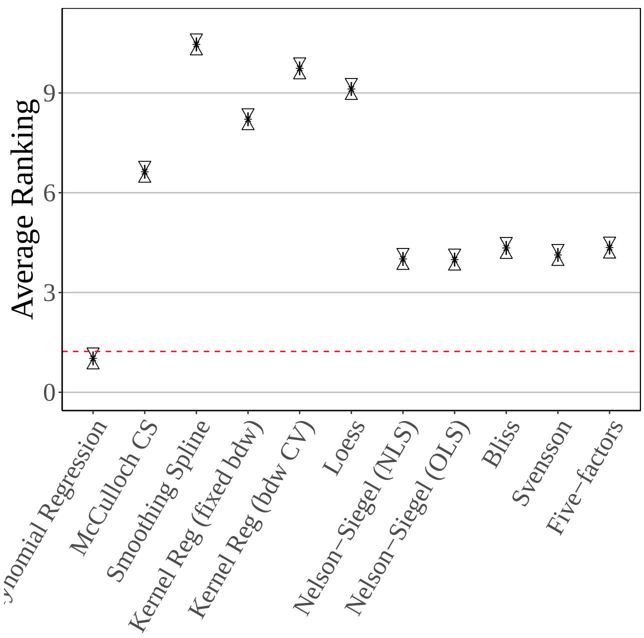
Figure 8: Multiple Comparison Procedure for Smoothness - Models Ranked by R, $R2$, $R3$



(a) R



(b) R_2



(c) R_3

Source: Elaborated by the author.

We analysed the two main spline specifications applied to yield curve estimation, McCulloch's Natural Cubic Spline and Smoothing Splines, and four specifications from the Nelson-Siegel family considering possible estimation methods for the baseline Nelson-Siegel model. As well that, we considered two Kernel regression specifications and the LOESS for estimating the Yield curve. This addition is a relevant contribution since the surveyed literature has ignored these options so far.

Using data from Brazilian interest rate derivatives over 1313 days, we compared model performance, evaluating them according to goodness-of-fit (in-sample and out-of-sample), robustness to outliers, and smoothness metrics. Besides the descriptive statistics on these metrics, the Friedman test and the multiple comparison procedure were used to assess the statistical significance of differences among the models. It's worth noting that the application of the multiple comparison procedure in the context of yield curve estimation appears to be a novel contribution.

As stated in the introduction, different applications of yield curve estimates may prefer some characteristics over others. The Smoothing Spline consistently has the best fit in-sample but is outperformed by the LOESS out-of-sample on all different maturity ranges. Among the Nelson-Siegel family models, the more parametrised versions have no clear advantage in terms of goodness-of-fit. However, the baseline model estimated by OLS under-performs the others. The robustness analysis shows that outliers mainly harm the Kernel regression estimates. At the same time, the Smoothing Spline and the LOESS are robust and have the best fit even in the presence of outliers. Finally, the smoothness analysis favours the parametric models. It also suggests that the smoothing spline is the worst option for extrapolating the yield for longer maturities.

References

Adams, K. J. and Deventer, D. R. V. (1994). Fitting Yield Curves and Forward Rate Curves With Maximum Smoothness. *The Journal of Fixed Income*, 4(1):52–62.

Anderson, N. and Sleath, J. (2001). New estimates of the UK real and nominal yield curves. Working Paper 11, Bank of England, London.

Annaert, J., Claes, A. G. P., De Ceuster, M. J. K., and Zhang, H. (2013). Estimating the spot rate curve using the Nelson–Siegel model: A ridge regression approach. *International Review of Economics & Finance*, 27:482–496.

Berger, P. L. (2015). *Mercado de Renda Fixa no Brasil - Ênfase em Títulos Públicos*. Editora Interciência, Rio de Janeiro, RJ.

Berk, R. A. (2016). *Statistical learning from a regression perspective*. Springer Berlin Heidelberg, New York, NY.

BIS (2005). Zero-Coupon Yield Curves: Technical Documentation. Technical report, Bank for International Settlements, Basel, Switzerland.

Björk, T. and Christensen, B. J. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance*, 9(4):323–348.

Bliss, R. R. (1996). Testing term structure estimation methods. Working Paper 96-12a, Federal Reserve Bank of Atlanta, Atlanta, GA.

Bolder, D. and Gusba, S. (2002). Exponentials, Polynomials, and Fourier Series: More Yield Curve Modelling at the Bank of Canada. Working Paper 2002-29, Bank of Canada.

Bolder, D. and Stréliski, D. (1999). Yield Curve Modelling at the Bank of Canada. Technical Reports 84, Bank of Canada.

Boor, C. D. (2001). *A Practical Guide to Splines*. Number 27 in Applied Mathematical Sciences. Springer Verlag, New York, revised edition.

Caldeira, J. F. (2011). Estimação da Estrutura a Termo da Curva de Juros no Brasil através de Modelos Paramétricos e Não Paramétricos. *Análise Econômica*, 29(55).

Caldeira, J. F., Moura, G. V., Santos, A. A. P., and Tourrucôo, F. (2016). Forecasting the yield curve with the arbitrage-free dynamic Nelson–Siegel model: Brazilian evidence. *EconomiA*, 17(2):221–237.

Chambers, D. R., Carleton, W. T., and Waldman, D. W. (1984). A New Approach to Estimation of the Term Structure of Interest Rates. *The Journal of Financial and Quantitative Analysis*, 19(3):233–252.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368):829–836. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1979.10481038.

Dybvig, P. H. and Ross, S. A. (1989). Arbitrage. In Eatwell, J., Milgate, M., and Newman, P., editors, *Finance*, The New Palgrave, pages 57–71. Palgrave Macmillan UK, London.

Estep, D. J. (2002). *Practical analysis in one variable*. Undergraduate texts in mathematics. Springer, New York.

Fisher, M., Nychka, D., and Zervos, D. (1995). Fitting the term structure of interest rates with smoothing splines. Technical Report 95-1, Board of Governors of the Federal Reserve System (U.S.).

Fraletti, P. B. (2004). *Ensaios sobre taxas de juros em reais e sua aplicação na análise financeira.* Tese de Doutorado (Administração), Universidade de São Paulo, São Paulo, SP.

Franklin Jr., S. L., Duarte, T. B., Neves, C. R., and Melo, E. F. L. (2012). A estrutura a termo de taxas de juros no Brasil: modelos, estimação e testes. *Economia Aplicada*, 16:255–290.

Gilli, M., Grosse, S., and Schumann, E. (2010). Calibrating the Nelson-Siegel-Svensson Model.

Gilli, M., Maringer, D., and Schumann, E. (2019). *Numerical methods and optimization in finance*. Academic Press is an imprint of Elsevier, London ; San Diego, CA, second edition edition.

Heath, D., Jarrow, R., and Morton, A. (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60(1):77–105.

Helwig, N. E. (2022). npreg: Nonparametric Regression via Smoothing Splines.

Hollander, M., Wolfe, D. A., and Chicken, E. (2014). *Nonparametric statistical methods*. Wiley series in probability and statistics. Wiley, Hoboken, NJ, third edition edition.

Howard, I. I. (2017). *Computational Methods for Numerical Analysis with R*. CRC Press, Boca Raton.

Hsu, J. (1996). *Multiple Comparisons - Theory and Methods*. CRC Press, Boca Raton, FL.

Hämmerlin, G. and Hoffman, K.-H. (1991). *Numerical Mathematics*. Undergraduate Texts in Mathematics. Springer, New York, NY.

Ioannides, M. (2003). A comparison of yield curve estimation techniques using UK data. *Journal of Banking & Finance*, 27(1):1–26.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer, New York NY.

Koning, A. J., Franses, P. H., Hibon, M., and Stekler, H. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting*, 21(3):397–409.

Li, Q. and Racine, J. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press, Princeton Oxford.

McCulloch, J. H. (1971). Measuring the Term Structure of Interest Rates. *The Journal of Business*, 44(1):19–31.

McCulloch, J. H. (1975). The Tax-Adjusted Yield Curve. *The Journal of Finance*, 30(3):811–830.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4):473–489.

Nymand-Andersen, P. (2018). Yield curve modelling and a conceptual framework for estimating yield curves: evidence from the European Central Bank's yield curves. Technical Report 27, European Central Bank.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer New York, New York, NY.

Rezende, R. B. and Ferreira, M. S. (2013). Modeling and Forecasting the Yield Curve by an Extended Nelson-Siegel Class of Models: A Quantile Autoregression Approach. *Journal of Forecasting*, 32(2):111–123.

Rezende, R. B. d. (2011). Flexibilizando os Modelos de Estrutura a Termo da Classe Nelson-Siegel. *Brazilian Review of Finance*, 9(1):27–49.

Seppälä, J. and Viertiö, P. (1996). The term structure of interest rates: Estimation and interpretation. Bank of Finland Research Discussion Paper 19/1996, Bank of Finland.

Svensson, L. E. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. Working Paper 4871, National Bureau of Economic Research.

Varga, G. (2009). Teste de Modelos Estatísticos para a Estrutura a Termo no Brasil. *Revista Brasileira de Economia*, 63(4).

Waggoner, D. (1997). Spline methods for extracting interest rate curves from coupon bond prices. FRB Atlanta Working Paper 97-10, Federal Reserve Bank of Atlanta, Atlanta, GA.

Wahlstrøm, R. R., Paraschiv, F., and Schürle, M. (2022). A Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Versions. *Computational Economics*, 59(3):967–1004.

Ypma, J., library), S. G. J. a. o. t. N. C., Borchers, H. W., Eddelbuettel, D., OS), B. R. b. p. o. m., OS), K. H. b. p. o. m., Chiquet, J., tests), A. A. r. d. c. f., Dai, X., Stamm, A., and Ooms, J. (2022). nloptr: R Interface to NLopt.