University of Brasília
Institute of Exact Sciences
Department of Statistics

Two essays on yield curve modelling

by

Matheus Stivali

Brasília, 12 December 2023

Two essays on yield curve modelling

by

Matheus Stivali

Dissertation submitted to the Department of Statistics at the University of Brasília, as part of the requirements required to obtain the Master Degree in Statistics.

Advisor: Prof. Dr. José Augusto Fiorucci

Brasília, 12 December 2023

Dissertation submitted to Graduate Program of the Department of Statistics at the University of Brasília, as part of the requirements required to obtain the Master Degree in Statistics.

Approved by:

Prof. Dr. José Augusto Fiorucci

Advisor, EST/UnB

Prof. Dr. Raul Yukihiro Matsushita

Examiner, EST/UNB

Prof. Dr. Geraldo Nunes Silva

Examiner, IBILCE/Unesp

"Police work wouldn't be possible without coffee", Wallander said.

"No work would be possible without coffee."

They pondered the importance of coffee in silence.

(Henning Mankell, *One Step Behind*)

Karma police, arrest this man

He talks in maths

(Radiohead, *Karma Police - OK Computer*)

To Olga.

Acknowledgments

I wish to convey my deepest appreciation to the faculty of PPGEST/UnB. In addition to their exceptional teaching, all the faculty members demonstrated unwavering support throughout the course, despite the challenges posed by the pandemic. My heartfelt gratitude goes to my advisor, Professor José Augusto Fiorucci, as well as to Professor Raul Matsushita and Professor Guilherme Souza Rodrigues.

I extend my special thanks to my colleagues at IPEA: Luis Henrique Paiva, André Rauen, and Luis Kubota. It would have been impossible to balance my work responsibilities and the demanding coursework without their invaluable assistance. Likewise, I would like to express my sincere appreciation to my former colleagues at the Labour Secretariat: Daniela, Tatiana, Luis Felipe (affectionately known as "BO" or "Miudinho"), and Bruno (nicknamed "Jovem" or "Interventor").

Lastly, I would be remiss not to mention my parents – who still inquire about my 'home-work' – and Carla and Olga, who stood by me throughout this journey. At the time, even one-year-old Olga played a role in my academic pursuits by assisting me in presentations for

the Sampling and Time Series disciplines.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo Expandido

DOIS ENSAIOS SOBRE A MODELAGEM DA CURVA DE JUROS

[Introdução] Entender o comportamento das taxas de juro é essencial para a gestão macroeconômica e para as decisões dos investidores privados. A taxa de juros de curto prazo é definida pela autoridade monetária de acordo com seus objetivos de política pública e essa taxa é obtida por meio de operações de mercado aberto. O comportamento das taxas de juros pagas para dívidas de prazo mais longo é influenciado pela taxa de curto prazo, mas esse é mais complexo e depende das expectativas em relação ao comportamento futuro das taxas de curto prazo e da inflação. A estrutura a termo das taxas de juros é a correspondência entre a maturidade de uma dívida (tempo até o vencimento) e o nível das taxas de juros associado a mesma, e sua representação gráfica é denominada curva de rendimentos. Esta pode assumir diferentes formas, a situação considerada normal é aquela em que as taxas de juros aumentam monotonamente com a maturidade. Curvas invertidas, em "forma de S" e *humped* ocorrem quando o mercado espera mudanças na taxa de curto prazo nos próximos meses ou anos. A dissertação avalia duas linhas de análise estatística da curva de juros para o Brasil: a primeira preocupada com a interpolação dos dados observados a cada dia para a estimação da curva completa, e a segunda preocupada com a extrapolação de informações passadas da curva de juros. Muitas das aplicações da curva de rendimentos dependem da relação entre maturidade e juros ser observável para todas as

maturidades, o que não ocorre. Em cada dia útil são observadas apenas alguns pontos da curva que correspondem aos títulos ou contratos futuros negociados naquele dia. Daí a relevância dos exercícios implementados no segundo capítulo em que várias técnicas de interpolação são utilizadas para obtenção da curva completa. Adicionalmente, a previsão da curva de rendimentos é uma ferramenta essencial para a estruturação da dívida pública, para a condução da política monetária, e para agentes privados que também emitem títulos de dívida ou compram os mesmos. Prever a curva de juros envolveria a modelagem das séries de juros de cada maturidade. Uma forma alternativa, mais parcimoniosa, foi proposta para Diebold e Li (2006). Tal abordagem é objeto do terceiro capítulo, comparando o desempenho desse modelo com técnicas de previsão de referência. **[Materiais e Métodos]** Para as análises foram utilizados dados de contratos futuros de taxas de juros (DI1) negociados no Brasil entre janeiro de 2018 e abril de 2023, totalizando 1313 dias úteis. Em cada dia são negociados em torno de 38 contratos de diferentes maturidades. O segundo capítulo desenvolve uma análise comparativa de técnicas de interpolação das taxas de juros que são estimadas em cada dia incluído na amostra. Os modelos abordados neste capítulo são chamados *empíricos*, pois não impõem restrições derivadas de modelos teóricos (econômicos) de estrutura de termo durante o processo de estimação. São considerados os modelos: regressão polinomial, modelos de *spline*, regressão de Kernel, regressão local (Loess), modelo Nelson-Siegel estimado por mínimos quadrados e mínimos quadrados não-lineares e extensões desse modelo (família Nelson-Siegel). Esses modelos são avaliados em relação a: qualidade do ajuste, robustez (em relação a *outliers*), e suavidade. Para a avaliação da qualidade do ajuste a cada dia é construído um conjunto de treinamento (*in-sample*) e um conjunto de validação (*out-of-sample*). A performance no conjunto de validação é o mais relevante para a avaliação dos modelos já que esse seria o problema típico subjacente a estimação de curva de rendimentos. Para a avaliação da robustez, a curva de rendimentos de cada dia da amostra é estimada duas vezes, uma com os dados originais e outra em que o nível da taxa de juros de uma maturidade selecionada aleatoriamente foi modificada por uma perturbação de mais ou menos, também definido aleatoriamente, 2%. Tanto para a avaliação da

qualidade do ajuste quanto para a robustez são utilizadas as métricas de Erro Quadrático Médio e Erro Médio Absoluto. Para a avaliação da suavidade são consideradas três métricas utilizadas na literatura baseadas na segunda derivada das funções estimadas. O terceiro capítulo se vale de estimativas dos parâmetros do modelo Nelson-Siegel feitas no segundo capítulo utilizando mínimos quadrados ordinários e mínimos quadrados não-lineares para implementar o modelo Diebold-Li. O filtro de Kalman é utilizado para avaliar a validade da interpretação dos parâmetros como variáveis latentes. As séries de estimativas são modeladas como três processos autorregressivos separados e como um vetor autorregressivo para fins de previsão. Os parâmetros preditos são então utilizados para estimar o nível de juros em maturidades específicas para avaliação da performance das previsões. Como modelos concorrentes são consideradas as previsões de *random-walk* e o modelo de suavização exponencial de Holt-Winters. Para avaliação da performance se utilizou a estratégia de "*walk-forward validation*", considerando um conjunto de treinamento inicial de 987 dias (75% da amostra). Destaca-se que o conjunto de validação (de 3 de janeiro de 2022 até 20 de abril de 2023) abarca um período de mudanças contínuas na forma da curva de rendimentos. Para a comparação dos diferentes modelos se utilizou o teste de Diebold-Mariano, com a modificação proposta por Harvey e outros. **[Resultados e Considerações Finais]** O segundo capítulo fez uma avaliação abrangente dos modelos de interpolação para estimar a curva de juros. Além dos modelos normalmente considerados pela literatura, foram considerados os modelos de regressão de Kernel e de regressão local (Loess) até então não aplicados a esse tipo de problema. Foram consideradas três dimensões para a comparação desses modelos, tanto na dimensão de qualidade do ajuste quanto na de robustez o modelo Loess apresentou o melhor desempenho fora da amostra sendo que em algumas situações ele não tinha um desempenho estatisticamente diferente do modelo de *smoothing splines*. Na dimensão relacionada a suavidade os modelos baseados em função (regressão polinomial e família Nelson-Siegel) tiveram o melhor desempenho. Para a comparação dos modelos foi utilizado o teste de comparações múltiplas, até então também não aplicado a esse tipo de problema. O terceiro capítulo fez uma implementação do modelo Diebold-Li (alternativamente chamado

de Nelson-Siegel dinâmico) para a economia brasileira recente. O modelo dinâmico de Nelson-Siegel teve um desempenho ruim em comparação com os resultados originais do Diebold and Li (2006) e alguns exercícios anteriores usando dados brasileiros de outros períodos. Em muitos casos, foi superado pela previsão de random-walk.

Palavras-chave: Estrutura de termo das taxas de juros, Curva de rendimentos, interpolação, splines, previsão, teste de comparações múltiplas, modelo Nelson-Siegel, modelo Diebold-Li.

Abstract

The dissertation undertakes two distinct lines of statistical analysis on the yield curve for Brazil: the first involves the interpolation of daily observed data to estimate the complete curve. In contrast, the second focuses on extrapolating past information to forecast the yield curve. These analyses aim to model the behaviour of interest rates in Brazil, offering insights for improved macroeconomic management and supporting investment decisions. The analysis utilizes data from interest rate futures contracts traded in Brazil between January 2018 and April 2023.

The second chapter is dedicated to estimating *empirical models* of the Term Structure of Interest Rates. Despite B3 periodically releasing yield curve estimates for monitoring the Brazilian market, various estimation techniques are considered for alternative purposes due to inherent trade-offs. The interest rate and maturity relationship holds for all terms, but daily observations are limited to specific maturities corresponding to traded securities or derivatives. Therefore, estimating the entire curve from these observed data points is crucial. This chapter evaluates *empirical* models, which do not impose restrictions derived from theoretical term structure models during the estimation process. These models are focused on obtaining a smooth function from observed data while adhering to specific constraints, such as the non-negativity of interest rates. The evaluation criteria include the quality of fit, robustness to outliers, and smoothness of the estimated function. This chapter contributes to literature by assessing models not previously applied to yield curve estimation and utilizing the multiple comparison procedure. Results highlight the strong fit of spline models, emphasize the greater smoothness of Nelson-Siegel family models, and recognize the noteworthy performance of the

previously overlooked Loess model.

The third chapter delves into modelling the yield curve dynamics through a factor model perspective to generate curve predictions. The analysis incorporates Brazilian data by implementing the Nelson-Siegel Dynamic model proposed by Diebold and Li (2006) and further developed in Diebold et al. (2006). Both original estimation procedures, two-step and one-step, are considered, focusing on the latter using the Kalman filter. Out-of-sample predictive capacity is assessed through the Diebold-Mariano test, comparing the performance of these implementations against simpler models.

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

Contents

List of Tables

List of Figures

Abbreviations and Acronyms

B3	B3 – Brazil Stock Exchange and Over-the-Counter Market
BCB	Brazilian Central Bank
B-Spline	Basis spline
CB	Coupon bond
CDF	Cumulative distribution function
Copom	Monetary Policy Committee
CNM	National Monetary Council
CDI	Interbank Certificate of Deposit
DI1	One-day Interbank Deposit Futures
DI rate	The average rate of interbank transactions carried out through CDI
DL	Diebold-Li model
GCV	Generalized Cross-Validation
MAE	Mean Absolute Error
MCP	Multiple Comparison Procedure
MSE	Mean Square Error
NS	Nelson-Siegel model
NLS	Nonlinear least squares
OLS	Ordinary least squares
PDF	Probability density function

Repo	Repurchase agreement
RMSE	Root mean square error
RSS	Residual sum of squares
Selic	Special System for Settlement and Custody
TSIR	Term structure of interest rates
SRF	Secretariat of Federal Revenue of Brazil
STN	National Treasury of Brazil
ZCB	Zero-coupon bond

List of Symbols and Notations

$\mathbb{1}_A(x)$	Indicator function: $\mathbb{1}_A(x) = 1$ if $x \in A$, and $\mathbb{1}_A(x) = 0$ otherwise.
$\beta_{0,t}$	coefficient, Eq. (2.18), associated with long-run interest rate
$\beta_{1,t}$	coefficient, Eq. (2.18), associated with short-run interest rate
$\beta_{2,t}$	coefficient, Eq. (2.18), associated with medium-run interest rate
λ	Tuning parameter for the smoothing splines; Bandwidth for the Kernel function; Decay parameter for the Nelson-Siegel family models
$b_k(.)$	Basis function
$b(m_i, k_j)$	Truncated Power Basis function
$B_{i,m}(.)$	B-Spline basis of order m
$d_t(m)$	Discount function evaluated at t for a maturity m
$f(t, t', T)$	Implied forward rate from t' to T evaluated at t
$K(.)$	Kernel function
m	Maturity
$y_t(m)$	Yield-to-maturity evaluated at t for a maturity m, yearly compounded
$\tilde{y}_t(m)$	Yield-to-maturity evaluated at t for a maturity m, continuously compounded
P_t	Market price of a bond evaluate at t
P_T	Face value (or Par value) of a bond
$R(.); R_2(.); R_3(.)$	Smoothness metrics
$\bar{R}_k$	Average ranking of model k

$r_{\alpha,K,D}$	Critical value for the multiple comparison procedure
S	Friedman rank sum test statistic
s	Span for the Loess

Chapter 1

Introduction

1.1 Initial Considerations

Understanding the behaviour of interest rates is crucial for macroeconomic management and private investors' decision-making. In most countries, the monetary authority establishes a reference interest rate based on its policy objectives, such as stabilising inflation or managing exchange rates. Subsequently, it utilises open market operations to achieve the reference rate. These operations encompass repurchase agreements ("repo") and reverse repurchase agreements ("reverse repo"), which involve the purchase or sale of a bond today with a commitment to sell or repurchase it the following day[1]. The slight price differential between the bought and sold bonds implicitly defines the overnight interest rate, thereby allowing the monetary authority to control short-term interest rates.

In Brazil, the National Monetary Council (*Conselho Monetário Nacional* - CNM) sets monetary policy objectives, including inflation targets for the next few years[2]. Based on the inflation

[1] To provide an accurate description of repo and reverse repo operations, it is worth noting that these transactions may involve spans of up to six months between the purchase and sale of the bond. However, one-day operations are the most common in the context of short-term interest rate determination.

[2] At present, the CNM consists of three members: the Finance Minister, the Planning Minister, and the Central Bank Governor. The CNM also has other responsibilities, such as determining the volume for various targeted credit policies, though these aspects are not pertinent to the present discussion.

target, the Brazilian Central Bank (BCB)'s Monetary Policy Committee (Copom)[3] convenes every 45 days to set the level of the Selic reference rate.

The term "Selic" stands for Special System for Settlement and Custody (*Sistema Especial de Liquidação e de Custódia*). It is a system developed and managed by the BCB for trading federal bonds in a dematerialised form. On a daily basis, as many financial agents engage in securities trading through Selic, the BCB conducts repo and reverse repo operations within the Selic system to maintain the interest rate around the level defined by the Copom.[4]

The Selic rate serves as the primary overnight rate in Brazil, although some overnight transactions may occur at slightly different rates. In addition to dealing with bonds at the BCB, banks perform short-term interbank borrowing operations using the Interbank Certificate of Deposit (*Certificado de Depósito Interbancário* - CDI). These interbank operations are intended to ensure that no bank ends a working day with either a negative cash balance or an excess of funds. The CDI rate represents the average rate of one-day interbank transactions conducted through interbank certificates of deposits. While the BCB calculates the Selic rate, B3 calculates the CDI rate.

As described earlier, the overnight interest rate is a product of daily transactions in the money market, governed by central bank interventions. The impact of changes in the overnight rate on long-term interest rates depends on how investors will adjust their portfolios. For instance, when the overnight rate falls, bonds with slightly longer maturities (e.g. one month) become more attractive as they offer higher interest rates for a modest increase in maturity. This interest rate differential leads to increased demand for these bonds, driving up their prices and reducing the interest rate for this maturity. This process subsequently recurs for longer maturities. Nevertheless, due to the imperfect substitutability of bonds with different maturities, the longer the maturity, the weaker the transmission of overnight rate changes. Consequently, **interest rates for longer terms** (ranging from some months to several decades) exhibit more

[3]The Copom is comprised of BCB's Board of Governors.

[4]The *Selic rate* is the average of rates used in all transactions conducted in the Selic system on a given day, while the *Selic reference rate* is the target set by the Copom and pursued by the BCB.

intricate behaviour and are influenced by market expectations regarding the future behaviour of overnight rates and inflation[5].

In the long run, various debt instruments exhibit distinct characteristics, including default risk, liquidity, taxation, and maturity, all of which affect the interest rates they bear. Default risk pertains to the possibility that a debt issuer may fail to meet its outstanding obligations, such as interest or the par value. For example, a company facing financial difficulties may cease its interest payments[6], and a financially distressed company may default on its debt altogether. In contrast, government bonds are generally considered **free of default risk**. Therefore, when assessing the appropriate interest rate for a bond with default risk, it is compared to the rate paid by government bonds. As a result, investors demand a higher interest rate for a bond issued by a private firm compared to what the government pays, and this disparity is known as the *risk premium*. The more uncertain the solvency of the debt issuer, the higher the risk premium.

Liquidity refers to the ease with which an asset can be converted into cash. Investors value liquidity because they may wish to pursue more profitable alternatives or rebalance their portfolios should the bond they hold start to depreciate. Consequently, a *liquidity premium* is established, with less liquid assets generally demanding higher interest rates. An asset's liquidity depends on the number of agents trading that asset and the frequency of transactions conducted by these agents in each period. More agents and transactions result in higher liquidity for an asset, and government bonds are typically the most liquid securities in the market.

Moreover, investors are concerned about the net (real) rate. Thus, the taxation of bond yields influences the interest rate required to invest in a bond. In Brazil, fixed-income instruments are subject to a standard taxation schedule, with rates declining from 22.5% (for maturities less than 180 days) to 15% (for maturities exceeding 720 days), with some exceptions[7]. Therefore,

[5] "[...] *the longer the relevant time horizon, the less can the authorities be regarded as in control.*", Goodhart (1989, p. 219).

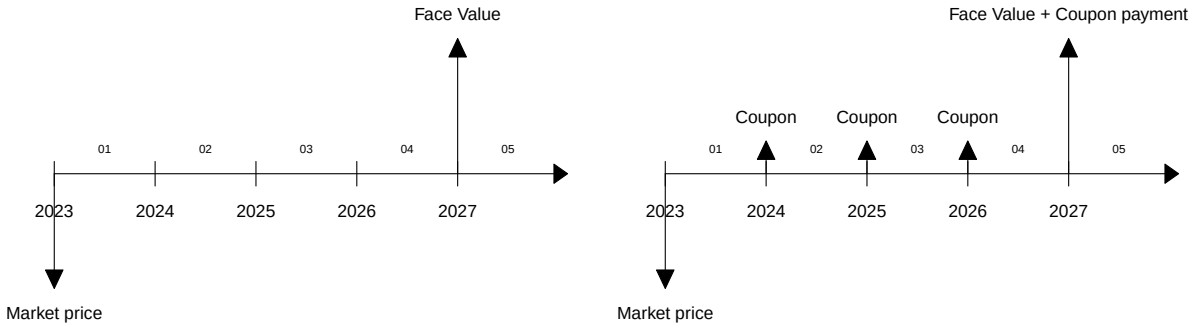
[6] As an example, in April 2023, the electricity distribution company "Light" (LIGT3), which operates in the state of Rio de Janeiro, suspended its debt payments, citing the need to rebalance its cash flow.

[7] An example of an exception is "*debentures incentivadas*," securities with special tax treatment intended to promote investment in specific sectors, such as infrastructure. Lower taxation for investors implies that the security issuer can offer a lower interest rate.

in most cases, government and private bonds face the same taxation[8].

Given the considerations regarding default risk, liquidity, and taxation, government bonds are the ideal reference for evaluating medium- and long-term interest rates. Maturity, or term to maturity, represents the time until a bond's principal is repaid. Depending on the bond's characteristics, it may pay interest only at maturity alongside the principal (referred to as a zero-coupon bond - ZCB), as depicted in Figure 1.1a, or it may provide intermediate interest payments (coupons), as shown in Figure 1.1b.

Figure 1.1: Cash flows for Zero-Coupon Bond and Coupon Bond



(a) ZCB Cash Flow

(b) Coupon Bond Cash Flow

Obtaining the interest rate for a given maturity is straightforward with ZCB data. One only needs to compare the repayment value (named *face* or *par value*, P_T) of a bond of the desired maturity (m) with its current market price (P_t). After that, adjust for the time to maturity ($m = T - t$) to obtain the interest rate per cent per year. This expression is called yield-to-maturity. For calculation convenience, yield-to-maturity is often considered as compounding interest continuously ($\tilde{y}_t$) instead of yearly compounding (y_t). There is a small but non-negligible difference between the two quantities, as demonstrated in the annex, illustrating how to derive one quantity from the other. The yield-to-maturity (y_t or $\tilde{y}_t$ depending on the

[8] Although there are generally no tax distinctions between government and private bonds, there is a difference in taxation based on the bond's maturity. Short-term investments are subject to higher tax rates compared to long-term investments. The regulation defines short-term investments as those with a redemption period of less than 365 days, while long-term investments have a redemption period exceeding this threshold (initially defined by *Instrução Normativa* SRF nº 487, 30 December 2004, and presently established in *Instrução Normativa* SRF nº 1585, 31 August 2015). This fact complicates comparisons of the net return on bonds with different maturities.

context) corresponds to the spot interest rate for a ZCB traded at time t, the reference date, that matures at time $T > t$:

$$\underbrace{y_t(m) = \left(\frac{P_T}{P_t}\right)^{1/m} - 1}_{\text{yearly compounded}} \qquad \underbrace{\tilde{y}_t(m) = \frac{ln\left(\frac{P_T}{P_t}\right)}{m}}_{\text{continuously compounded}}.$$

The **term structure of interest rates** is the mapping from time to maturity to ZCB interest rates level: 3 months $\mapsto 2\%$ per annum, 6 months $\mapsto 2.5\%$ per annum. Therefore, since maturities and interest rates are always positive, one can write:

$$y_t : \mathbb{R}_0^+ \to \mathbb{R}_0^+.$$

The subscript *t* indicates that the term structure of interest rates evolves; each working day has a different yield curve. Besides the **yield-to-maturity** (or spot rate) form, other related concepts may express the term structure of interest rates. Instead, one could consider the **discount function**:

$$d_t(m) = e^{-\tilde{y}_t(m)\times m}.$$

Considering a ZCB with face value of 1 monetary unit and maturity of m, $d_t(m)$ represents the discount applied on the bond's face value ($P_T = 1$) such that it matches the bond's current market price (P_t). Otherwise, one could regard the **implied forward rate** $f(t, t', T)$, which is obtained from two ZCB with the same reference date (t), but different maturity dates (t' and T, with $t' < T$). Considering interest continuously compounded ($\tilde{y}_t$), one may write:

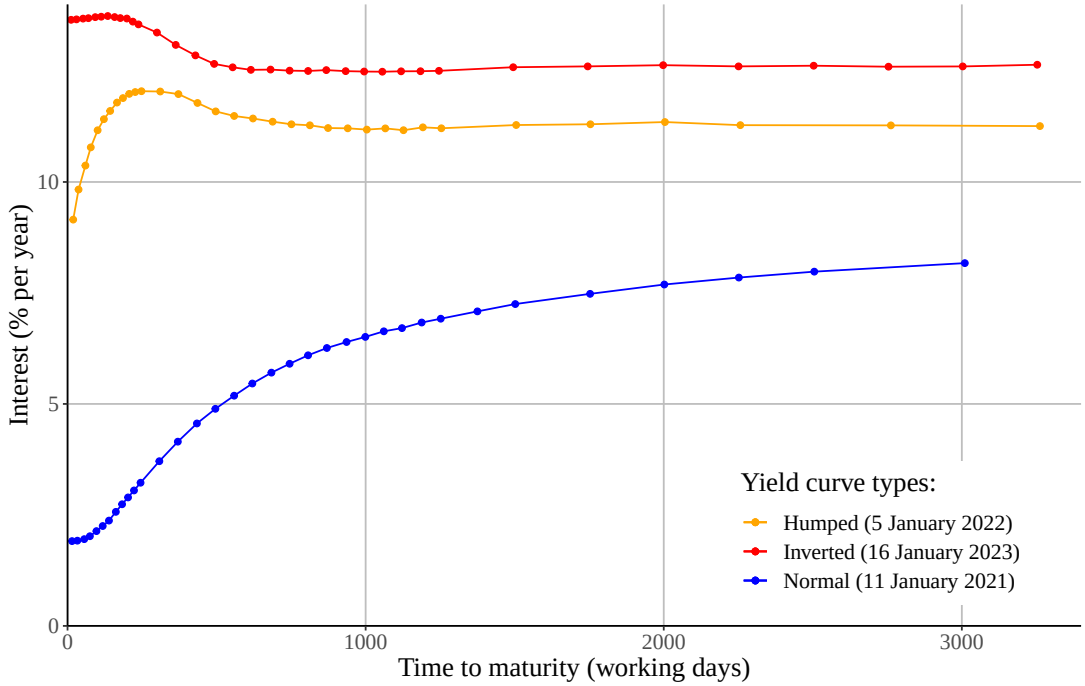
$$f(t, t', T) = \frac{(T-t)\times\tilde{y}_t(m) - (t'-t)\times\tilde{y}_t(t')}{(T-t')}.$$

The forward rate represents the rate for a debt starting in t' and maturing in T *implied* by

bond prices in t.

Interest rates in this context are often called yield, and the graphic representation of the term structure of interest rates is called the **yield curve**. As shown in Figure 1.2, yield curve can take different forms. The case in which interest rates monotonically increase with the time to maturity is considered the "normal" scenario, with inverted, "S-shaped", and "humped" curves occurring when the market expects changes in the short-run rate in the following months or years.

Figure 1.2: Yield Curve shapes



Source: Elaborated by the author using DI1 data from B3.

Besides the different shapes that yield curves can have, Mishkin (2000, p. 92-93) and Diebold and Li (2006, p. 343) mention some stylised facts that should be taken into account when modelling it:

- Typically, yield curves exhibit an upward slope (positive inclination) and concavity. Therefore, this configuration is considered the normal shape for a yield curve.

- Commonly, the short-term rate governs the form of the yield curve. If short-term rates are low, one would expect a positive inclination. Conversely, when short-term rates are high, one would expect a flat curve or a negative inclination (i.e. an inverted yield curve).
  - The short end of the yield curve is more volatile than the long end.
  - Long rates are more persistent than short rates.
- Interest rates for bonds with different maturities move together over time, i.e., a change in the short-run rate happens concurrently with changes in the same direction (yet with different magnitudes) by long-run rates.
- Yield dynamics are persistent, and spread[9] dynamics are much less persistent.

Theories of the term structure

Various theories have attempted to explain the economic rationale behind yield curve shapes and the empirical regularities mentioned earlier. However, no single theory comprehensively accounts for all observed yield curve behaviours, and much of the empirical research on the term structure pertains to testing these theories. Nonetheless, one can provide a concise overview of the leading theories as follows:

- **Expectation hypothesis**: The expectation hypothesis posits that the interest rate for a long-term debt (e.g., maturing in 10 years) equates to the average of short-term rates (monthly or yearly) expected over this period. According to this theory, an investor would be indifferent between purchasing and holding a ten-year bond until maturity or buying and reinvesting in a one-year bond annually over ten years. Following Campbell and Shiller (1991), the expectation hypothesis may be formally expressed as:

$$y_t(n) = \left(\frac{1}{k} \sum_{i=0}^{k-1} E_t \left[y_{t+mi}(m) \right] \right) + c, \quad k = n/m \in \mathbb{Z}. \tag{1.1}$$

[9]The spread between the n-period rate and the m-period rate, with $n > m$ is: $s_t(n, m) = y_t(n) - y_t(m)$.

Here, $y_t(n)$ represents the yearly interest paid for a bond maturing in n years at period t, and $y_{t+mi}(m)$ is the yearly interest for a shorter bond in subsequent periods. The term c denotes an excess return that an n-period bond offers over an m-period bond and is referred to as the *term premium*. This term premium may vary based on the difference between these maturities $(n - m)$ but is considered constant over time. In the context of the pure expectation hypothesis, assuming $c = 0$, the explanation for a yield curve with a normal shape would entail agents expecting short-term interest rates to rise in the forthcoming periods.[10]

  - *Fisher effect*: The Fisher effect pertains to the tendency for an increase in expected inflation to lead to a rise in nominal interest rates, first proposed by Fisher (1930). It can be approximated as $i \approx r + E_t[\pi]$, which is an approximation of $(1 + i) = (1 + r)(1 + E_t[\pi])$, where i signifies the nominal interest rate, r represents the real interest rate, and $E_t[\pi]$ denotes the expected inflation at period t. Utilising the Fisher equation, it is possible to infer the market's expectations about inflation by comparing the yield-to-maturity of inflation-indexed and nominal bonds (Valentim, 2022).

- **Liquidity preference**: The liquidity preference hypothesis asserts that investors prefer to convert their investments into cash sooner, demonstrating a preference for liquidity. Consequently, they would demand a higher yield – a liquidity premium – to hold long-term bonds rather than short-term bonds. This liquidity premium accounts for the *inflation and interest rate risks* associated with long-term bonds. If inflation or overnight interest rates rise, the value of a long-term bond would be more affected than that of a short-term bond. This hypothesis was initially proposed by Hicks (1946) and evolved from Keynes' liquidity preference theory. Formally, the liquidity preference hypothesis implies $c \neq 0$ in equation 1.1.

[10] Explanations for the term structure following the expectation hypothesis date back to Fisher (1896).

- **Segmented market**: The segmented market hypothesis argues that bonds with different maturities are not substitutes; supply and demand at each maturity level set the interest rate level. Consequently, distinct maturity ranges correspond to different bond markets. For instance, while banks focus their transactions on short-term bonds due to their preference for liquidity, pension funds and insurance companies strongly prefer long-term bonds. An extreme version of this hypothesis would consider expectations about future rates as irrelevant. This theory was introduced by Culbertson (1957).

- **Preferred habitat**: The preferred habitat hypothesis contrasts with the segmented market hypothesis because it assumes that bonds with different maturities are substitutes, albeit imperfect ones. Therefore, an investor would purchase a bond with a maturity outside its preferred range (the "*preferred habitat*") if it offers a premium. However, the preferred habitat hypothesis has the same consequence as the liquidity preference hypothesis, meaning that equation 1.1 would have $c \neq 0$. The preferred habitat hypothesis was put forward by Modigliani and Sutch (1966, 1969).

1.2 General Objective

The principal objective of this dissertation is to model the behaviour of interest rates in Brazil, with the aim of enhancing macroeconomic management and providing support for investment decisions. To accomplish this objective, we pursue two interconnected approaches, as outlined below:

Chapter 2 assesses empirical models of the Term Structure of Interest Rates. Although B3 periodically releases yield curve estimates to monitor the Brazilian market (B3 Manual, 2022), we explore various estimation techniques. Each technique yields estimates with distinct characteristics, involving a trade-off between goodness-of-fit and smoothness. Different applications may prioritise specific estimate features based on their requirements. Thus, we map how well each technique performs along distinct dimensions.

Although the relationship between yield and maturity applies to all maturities, each day, we observe only a few data points corresponding to the bonds or derivatives traded on that working day. Therefore, it is necessary to estimate the entire curve based on these data points. The models examined in this chapter are referred to as empirical models because they do not impose constraints from term structure theoretical models during the estimation process. These models are solely concerned with deriving a smooth function from observed data points while adhering to certain restrictions, such as ensuring non-negativity of interest rates. The evaluation of these models encompasses criteria such as goodness-of-fit, stability/robustness to outliers, and smoothness. These empirical models differ from equilibrium models, which are not considered in this dissertation, as they derive specific functional forms from general equilibrium models before estimation.

The main goal of Chapter 2 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 in a similar vein to this chapter, such as Varga (2009) and Caldeira (2011), are around a decade old. The recent fiscal and monetary developments in Brazil may have altered the conclusions of these earlier analyses. Additionally, these previous studies encompass a different range of models compared to those examined here. Therefore, Chapter 2 offers an up-to-date evaluation of yield curve estimation, providing valuable support for asset pricing and macroeconomic analysis.

Chapter 3 adopts a factor model perspective to delve into the dynamics of the yield curve, aiming to generate predictions. These models posit that a small number of factors can effectively summarise the entire yield curve, which comprises interest rates at various maturities on a given day. Analysing the dynamics of these factors provides insights into how the yield curve evolves. The concept of factor models for yield curve analysis dates back to the work of Litterman and Scheinkman (1991). Notably, the parameters from the Nelson-Siegel model, explored in Chapter 2, can be interpreted as factors associated with yield curve's level, steepness, and curvature.

The analysis specifically incorporates Brazilian data by implementing the Nelson-Siegel Dynamic model proposed by Diebold and Li (2006) and developed in Diebold et al. (2006). Both original estimation procedures, two-step and one-step, are considered, focusing on the latter utilising the Kalman filter. The out-of-sample predictive capacity is rigorously assessed through the Diebold-Mariano test, allowing for a robust comparison of the performance of these implementations against simpler models.

1.3 Data Sources

As previously discussed, 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 proprietary databases. For example, Anbima provides data on the transactions of the last five working days for free.[11] For more extended periods, the data is exclusively available through subscription, limiting access to researchers with institutional database access.[12]

An alternative data source for federal government bonds is the Tesouro Direto system[13], developed by STN, enabling individuals to directly trade federal bonds with the Treasury. However, this database provides a partial market perspective, capturing only transactions between individuals and STN, excluding banks and institutional investors. Furthermore, during periods of heightened bond market volatility, STN may suspend transactions, diminishing the research utility of these data.

As a result, most empirical studies on the term structure of interest rates turn to data from One-day Interbank Deposit Futures ("DI1")[14] contracts provided by B3[15]. These data represent

[11] Ambima - taxas de títulos públicos.

[12] The work by Souza Junior (2021) uses this data, for instance.

[13] STN - Tesouro Direto.

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

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

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. In some cases, researchers may opt to include the first point (maturity = 0) in the term structure, usually derived from the CDI-Overnight rate (Interbank Certificate of Deposit Overnight).[16]

[16]Over time, the Selic and CDI-Over rates tend to move in tandem, but the CDI rate consistently hovers slightly below. According to IMF (2018), the Selic rate provides a more accurate reflection of the true capital cost as it is more representative of the market. It encompasses all repo market transactions among participants in the SELIC system. Furthermore, the Selic market boasts significantly greater magnitude compared to the CDI market. On the other hand, the CDI rate considers only a subset of interbank transactions, specifically those at fixed interest rates, excluding the ones involving floating overnight DI transactions. The marginal disparity in the CDI rate compared to Selic can be attributed to the dominant market influence of larger banks, as suggested by the IMF. These banks possess greater liquidity and hold an advantage over smaller banks with surplus funds but limited investment options. Fraletti (2004, 54-56) elaborates on the disparity between these rates, citing the lack of liquidity in the CDI-Over market as a contributing factor. However, the author emphasises that CDI-Over tends to converge with Selic as arbitrage opportunities diminish, and for most financial applications, both rates are considered as substitutes.

Chapter 2

Estimating the Term Structure of Interest Rates: a comparison of techniques using Brazilian data

2.1 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 it's 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. Thus, estimating the term structure from the available data points is essential. 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. Additionally, a yield curve estimate should ideally exhibit key economic characteristics, including non-negativity of interest rates and an upper limit on long-term rates.

This chapter employs various techniques to estimate the Term Structure of Interest Rates using Brazilian data spanning from January 2018 to April 2023. This timeframe encompasses diverse economic scenarios that give rise to various forms of the yield curve, as reported in literature. The comparison of these techniques considers their goodness-of-fit, robustness to outliers, and smoothness.

The choice of 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. The objective here isn't to rank these techniques but to comprehensively study their strengths and weaknesses.

As outlined in the previous chapter, three interrelated concepts allow us to describe the term structure: yield-to-maturity, discount function, and forward rates. Given that the data used in this chapter is already in a yield-to-maturity format, this concept forms the foundation for all modelling exercises.

This chapter is structured into five sections. Section 2.2 delves into spline and function-based techniques, providing a review of existing literature that has applied these methods to yield curve estimation. Section 2.3 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 2.4 offers insights into the data employed and outlines the estimation procedures for implementing the presented models. Section 2.5 unveils the results, while Section 2.6 concludes this chapter.

2.2 Estimation methods of the term structure of interest rates

Estimating a nonlinear relationship, which may have different shapes, is a challenge commonly encountered in various statistical applications. From a limited dataset, the objective is to discern a nonlinear function denoted as $y(m) = f(m) + \epsilon$, which governs the underlying data.

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.

To illustrate the flexibility of the techniques, we will consider data from two yield curves with different shapes in this section: a normal curve, using data from 22 October 2018, and an inverted curve, with data from 16 January 2023.

2.2.1 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 $\xi = \{k_0, k_1, \ldots, k_K\}$, which partitions a given interval $[a, b] \subset \mathbb{R}$ into subintervals is called a *knots sequence* (or a *knot set*), where $a = k_0 < k_1 < \ldots < k_K = b$. The points $k_1, \ldots, k_{K-1}$ 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.

[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.

This subsection presents different splines specifications and how they have been used on the yield curve estimation.

Linear interpolation

A first approach to model a nonlinear relationship can be to approximate it by a set of different 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:

$$y_t(m) = \beta_0 + \beta_1 m + \beta_2(m-k_1)\mathbb{1}_{[k_1,k_2[}(m) + \beta_3(m-k_2)\mathbb{1}_{[k_2,\infty[}(m) + \epsilon_{t,m}. \tag{2.1}$$

A similar alternative would be to model the nonlinear relationship using *step functions*, which corresponds to converting a continuous variable into an ordered categorical variable, (James et al., 2021, p. 290). For that, one would create K knots ($k_1, k_2, \ldots, k_K$) in the relevant range and then construct K new variables using indicator functions: $\mathbb{1}_{[k_1,k_2[}(m)$, $\mathbb{1}_{[k_2,k_3[}(m)$, ..., $\mathbb{1}_{[k_K,\infty[}(m)$. In this case, the relevant regression is $y_t(m) = \beta_0 + \beta_1\mathbb{1}_{[k_1,k_2[}(m) + \beta_2\mathbb{1}_{[k_2,k_3[}(m) + \ldots + \beta_K\mathbb{1}_{[k_{K-1},k_K[}(m) + \epsilon_{t,m}$.

While the estimation of equation (2.1) is a simple way to estimate the yield curve, the resulting curve is not smooth, and there are no derivatives at the knots, (Valentim, 2022, p. 89). Therefore this approach is inappropriate for many applications that demand differentiability from the yield curve estimates.

The choice of the number of knots in linear interpolation illustrates the bias-variance trade-off, which also is relevant in more complex splines implementations. On the one hand, increas-

ing the number of knots will provide a better fit, and, on the limit, one could consider each observation as a knot (maximising bias and minimising variance). On the other hand, more knots will imply a loss of degrees of freedom and increase the number of non-differentiable points.

Polynomial regression

A more direct approach to model a nonlinear relationship is to engage with **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 (2.2) and estimated by least squares as an usual linear model, (James et al., 2021, p. 290):

$$y_t(m) = \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 m^3 + \ldots + \beta_N m^N + \epsilon_{t,m}. \tag{2.2}$$

However, considering polynomials of higher degree, the estimation procedure is troublesome. Consider the matrix representation of the linear model $y = \mathbf{X}\beta$:

[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).

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{N+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & m_1 & m_1^2 & m_1^3 & \dots & m_1^N \\ 1 & m_2 & m_2^2 & m_2^3 & \dots & m_2^N \\ 1 & m_3 & m_3^2 & m_3^3 & \dots & m_3^N \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & m_{N+1} & m_{N+1}^2 & m_{N+1}^3 & \dots & m_{N+1}^N \end{bmatrix}}_{\mathbf{X}=\text{Vandermonde matrix}} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{N+1} \end{bmatrix} .$$

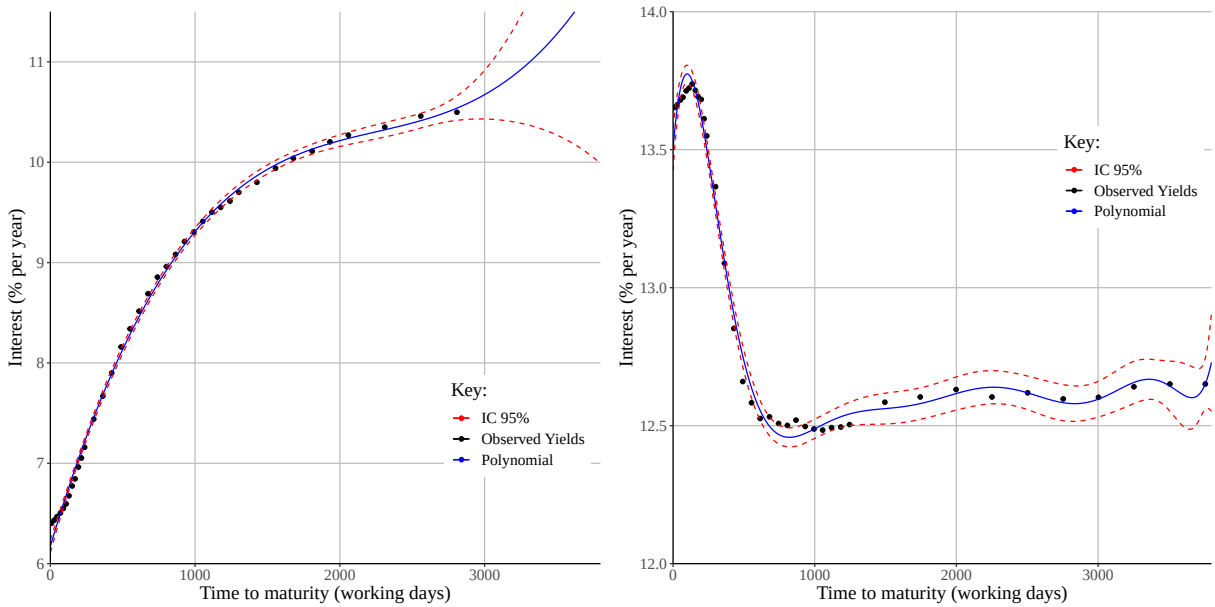
Determining the coefficients β by least squares depends on inverting the *Vandermonde matrix*, which is plagued by numerical problems (Bolder and Gusba, 2002, p. 4-6; Howard, 2017, p. 97-99). Besides the computational problems, a higher degree polynomial regression has other issues. For example, although 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 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.

Evidence supporting this conclusion can be seen in Figure (2.1), which presents two polynomial regressions applied to the Brazilian data. Whereas a polynomial of degree 4 was enough to fit a normal curve (Figure 2.1a), the inverted curve, which has a more complex form at shorter maturities, demanded a higher degree polynomial (degree 9) to attain a satisfactory fit (Figure 2.1b). Notice that the extrapolated estimates at the long range have a behaviour inconsistent with the economic theory. An unbounded growth at longer maturities can be inferred from

Figure (2.1a).

Figure 2.1: Polynomial regression



(a) Polynomial regression - degree 4

(b) Polynomial regression - degree 9

Source: Elaborated by the author.

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 (2.2) 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 (2.2) and used ordinary least squares.

Regression splines

The approaches presented previously 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 $b_1(m_i)$, $b_2(m_i)$, ..., $b_K(m_i)$:

$$y_i = \beta_0 + \beta_1 b_1(m_i) + \beta_2 b_2(m_i) + \beta_3 b_2(m_i) + \ldots + \beta_K b_K(m_i) + \epsilon_i. \tag{2.3}$$

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 (2.4), using least squares for each data subset delimited by the knots:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i \qquad x_i \in [k_j, k_{j+1}[. \tag{2.4}$$

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 (2.4) 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:

$$y_i = \beta_0 + \beta_1 m_i + \beta_2 m_i^2 + \beta_3 m_i^3 + \sum_{j=1}^{K} \theta_j (m-k)_+^r + \epsilon_i \tag{2.5}$$

$$= \beta_0 b_1(X) + \beta_1 b_2(X) + \beta_2 b_3(X) + \beta_3 b_4(X) + \sum_{j=1}^{K} \theta_j b(m_i, k_j) + \epsilon_i. \tag{2.6}$$

Hence, a basis representing a cubic spline with three knots (ξ = {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 (2.5) can be estimated by least squares. Considering $K = 3$, three knots divide the predictor's range into four subsets. In each subset, one needs to estimate four parameters (one intercept and three coefficients). Thus estimating a cubic spline with 3 knots corresponds to estimating $3 \times 4 = 12$ regression coefficients. The cubic spline imposes three continuity

[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).

conditions in each knot, which frees $3 \times 3 = 9$ degrees of freedom. Thus estimating a cubic spline with three knots demands $12 - 9 = 3$ degrees of freedom.

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.

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.

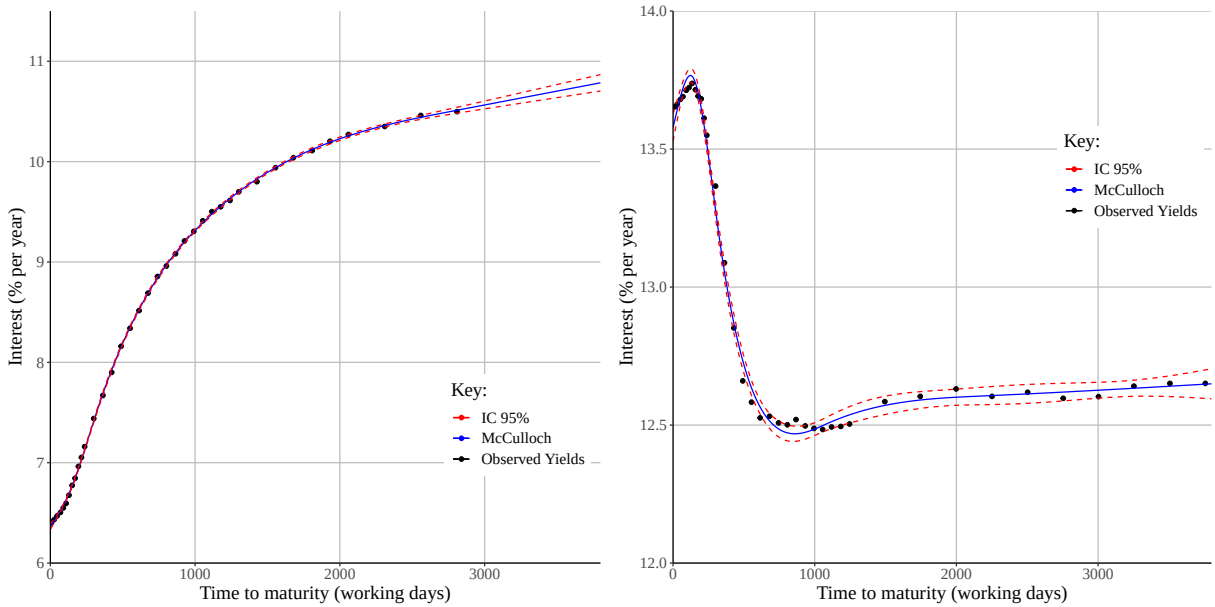
McCulloch (1971) expressed the problem of choosing the appropriate number of knots and their positions regarding the bias-variance trade-off[4] even though he did not use this terminology. McCulloch suggested that one should choose the number of knots to minimise the unbiased estimator of the variance. However, because this is a computationally expensive method with more than one local minimum, this author suggests determining the number of knots (K)

[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).

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 $n/K(n)$).

Two implementations of natural cubic splines following McCulloch's rule for knot quantity and position using Brazilian data can be seen in Figure (2.2). The extrapolated estimates are well-behaved, not showing abrupt curve changes after the last observation, which contrasts with polynomial regression estimates.

Figure 2.2: McCulloch Cubic Splines



(a) Normal Yield Curve

(b) Inverted Yield Curve

Source: Elaborated by the author.

Exponential splines

Vasicek and Fong (1982) proposed the *exponential splines* to model the discount curve as a superior alternative to McCulloch's cubic splines implementations. According to these authors, piecewise polynomials would not be an appropriate choice to model the discount curve because this is a known exponential decay relation. The local fit of a piecewise polynomial could not be

good when approximating such functions, despite its flexibility.

Considering that the discount function has the form $d_t(m) = e^{-\gamma \times m}$, they suggested a variable transformation such that the exponential function was transformed into a power function, which polynomial splines could better approximate. Considering an arbitrary constant α, the transformed variable x, with $0 \leq x < 1$, would be obtained in the following way:

$$m = -\frac{1}{\alpha} ln(1-x) \Rightarrow d_t(m) = e^{-\gamma \times m} \Rightarrow d_t(m) = d_t\left(-\frac{1}{\alpha} ln(1-x)\right) = g_t(x)$$
$$g_t(x) = e^{-\gamma \times -\frac{1}{\alpha} ln(1-x))} = e^{\frac{\gamma}{\alpha} ln(1-x)} = (1-x)^{\frac{\gamma}{\alpha}}.$$

According to Vasicek and Fong (1982), the problems with polynomial splines are that they "weave" around the underlying exponential function points, resulting in a poor local fit, and would be incapable of reproducing the exponential decay at longer maturities. Therefore, approximating $g_t(x) = (1-x)^{\frac{\gamma}{\alpha}}$ by splines would surpass these faults. However, Shea (1985) pointed out that while Vasicek and Fong (1982) were right about the difficulties in approximating an exponential function by a polynomial, that was not the case when one considers local polynomial approximations, as splines. Furthermore, Shea concludes that estimates from *exponential splines* give estimates identical to those obtained by polynomial splines but demand a more complex procedure.

Since the data used in this dissertation allows modelling yield-to-maturity directly, avoiding discount curve estimation, we will not consider exponential splines. Furthermore, Steeley (2008) found that superior yield curve estimates are obtained when the yield curve is fitted directly instead of fitting the discount function and then obtaining the yield curve.

B-Splines representation

According to Boor (2001), the collection of all piecewise polynomial functions of order M for a given set of K breaks corresponds to a linear space of dimension MK. Therefore, the piecewise polynomial function (f) that approximates an underlying unknown function (that generated the

data) can be obtained from a basis for that linear space. Furthermore, when one imposes some conditions of *f*, such as first and second derivatives continuity, it is possible to consider the subspace corresponding to those conditions.

These facts make it possible to express any spline as a linear combination of the proper basis. The basis for the space of all splines is called basis splines (**B-Splines**) and Boor (2001, 87-103) presents a method to obtain it. Such basis depends on the spline degree and on the number of knots (breaks). To obtain the B-Spline basis function of order m, $B_{i,m}(x)$, one must expand the original sequence of K breaks, corresponding to the knots sequence $k_0 < k_1 < \ldots < k_K$. This expansion is done by adding M new knots before k_0 and after k_K, with $m \leq M$:

$$\underbrace{\tau_{-M} \leq \ldots \leq \tau_{-1} \leq \underbrace{k_0 < \ldots < k_K}_{\xi=\text{Original knots sequence}} \leq \tau_{K+1} \leq \ldots \leq \tau_{K+M}}_{\tau=\text{Expanded knots sequence}} .$$

The values defined for τ_i are arbitrary and may be the same as the boundary knots (k_0 for $i < 0$ and k_K for $i > K$). Despite this redundancy, the additional knots are necessary due to the recursive construction of higher-order B-Splines. Then, one can apply equation (2.7), corresponding to the *Haar basis functions*, to the expanded knots sequence to obtain first-order B-Splines. After that, higher order B-Splines basis, say of order n, can be calculated using equation (2.8) and values from the previous order $(n-1)$ as follows:

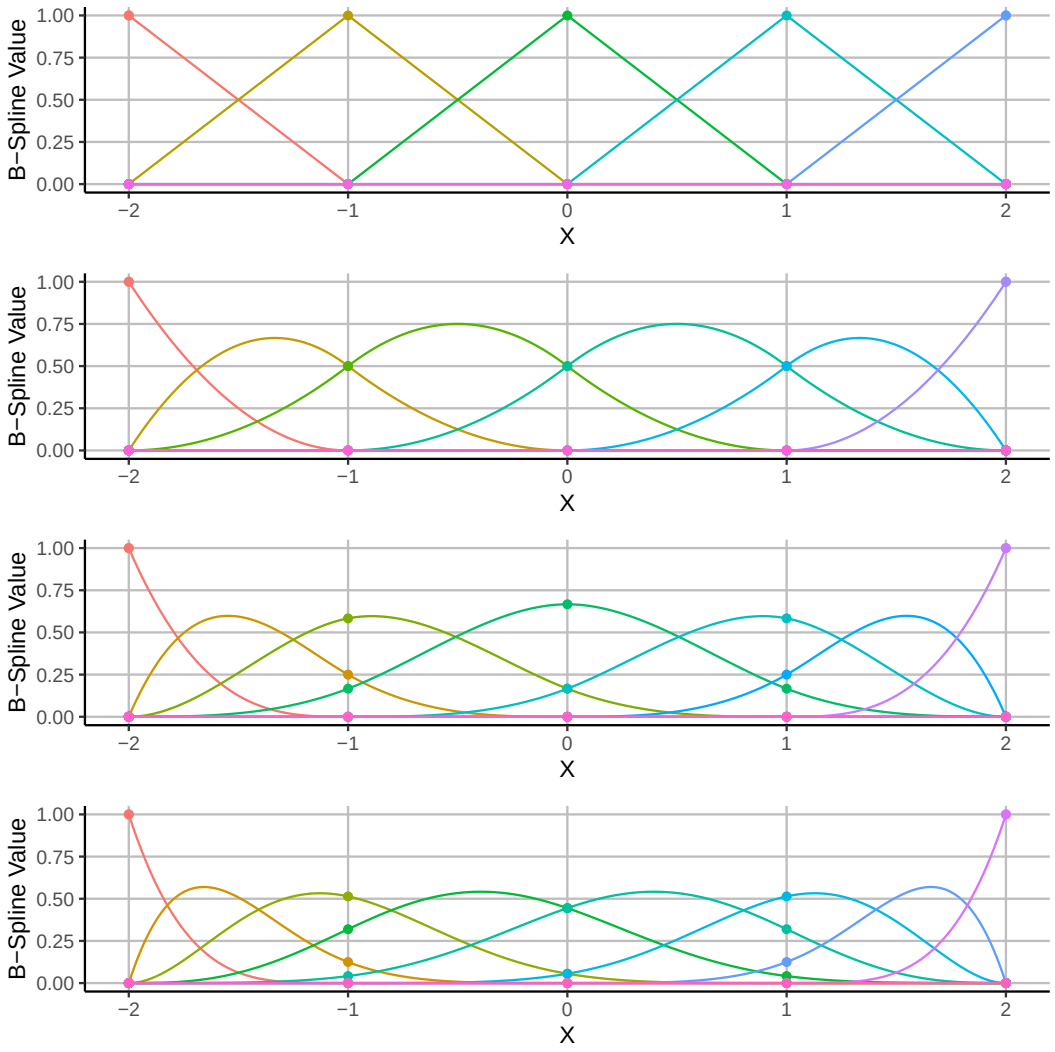
$$B_{i,1}(x) = \begin{cases} 1 & \text{if } \tau_i \leq x < \tau_{i+1} \\ 0 & \text{otherwise} \end{cases} \qquad \text{for i} = -M, \ldots, M \tag{2.7}$$

$$B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+m} - x}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(x) \qquad \text{for } m > 1. \tag{2.8}$$

This recursive basis construction will produce sets of functions for each order, the B-Spline basis of that order, over the ranges delimited by the knots. Figure 2.3 panels show some examples; each colour represents a different function on each order B-Spline basis. The number

of functions on a B-Spline basis equals the spline order (m) plus the number of (internal) knots (K). For instance, consider Figure 2.3 third panel, that shows a cubic B-Spline ($m = M = 4$) basis over $[-2, 2]$ with three *interior* knots (the sequence $\xi = -1, 0, 1$), thus $K = 3$). In this case, one has a B-Spline basis of order 4, composed of seven basis functions ($m + K = 7$). There are four non-zero basis functions for each segment delimited by adjacent knots. Adjacent segments, e.g. $[-1, 0]$ and $[0, 1]$, have three common non-zero basis functions.

Figure 2.3: B-Spline basis of orders 1 to 4, with interior knots at {-1,0,1}



Source: Elaborated by the author.

Concluding, for a spline of order m with K knots, one would have a B-Spline basis with $m + K$ functions. The spline can then be expressed as a linear combination of these basis functions ($B_{i,m}(x)$, $i = 1, \ldots, m+K$), which is called the *B-form*. However, $B_{1,m}(x)$ is usually dropped to avoid perfect correlation when estimating a spline represented in the B-form. Thus it is considered as the intercept. In contrast, the other terms will translate into columns on the predictor matrix

$$f(x) = \sum_{i=1}^{m+K} \beta_i B_{i,m}(x) \qquad x \in [k_0, k_K]. \tag{2.9}$$

As pointed out by Berk (2016, p. 66-69), B-Spline representation is a computationally convenient tool to construct piecewise cubic polynomials and natural cubic splines; there is no substantive meaning in the β coefficients in the equation (2.9). However, Hastie et al. (2009, p. 187-189) emphasize that many knots in a large sample may have an unbearable computational cost.

Since most applications of B-Splines for the term structure estimation consider it with smoothing splines instead of regression splines we review this literature on the following session.

Smoothing splines

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(x) will not be smooth. Therefore, regularisation is necessary to balance fit quality and smoothness. This can be done considering equation (2.10), in which $\lambda \geq 0$ is a *tuning parameter*

$$RSS(f, \lambda) = \sum_{i=1}^{N} [y_i - f(x_i)]^2 + \lambda \int [f''(t)]^2 \, dt. \tag{2.10}$$

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 $f(.)$ second derivative at each data point and corresponds to a measure of the roughness of $f(.)$.

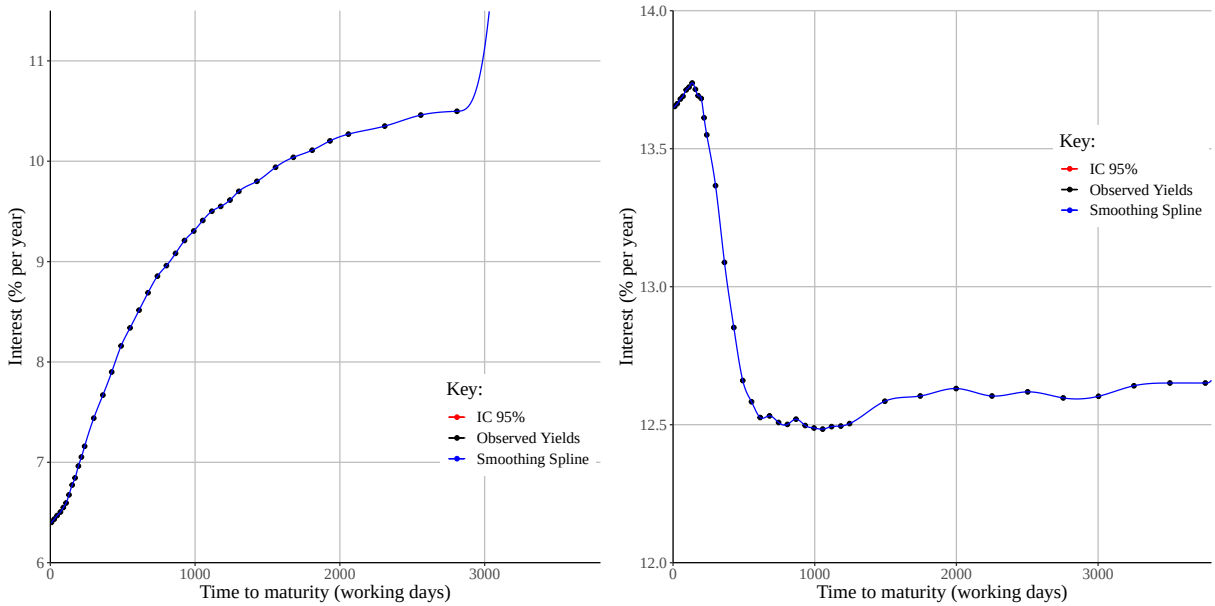
James et al. (2021, p. 302) point out that the minimisation of equation (2.10) results in a shrunken version of the natural cubic spline. The tuning parameter λ adjusts the bias-variance trade-off. If $\lambda = 0$, the penalty term will not affect the estimation, interpolating all points. On the other hand, a high value for λ 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 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). The effects of different λ values are illustrated in Figures 2.4, 2.5, and 2.6.

Regarding the implementation that determined the λ value by cross-validation, one can see that it does not appear to overfit the data and has narrow confidence bands. However, when ex-

amining the extrapolated values for the normal curve in Figure (2.5a), a behaviour inconsistent with the economic theory becomes apparent, as it displays a slight decreasing trend beyond the last observation.

Figure 2.4: Smoothing splines - $\lambda = 0$



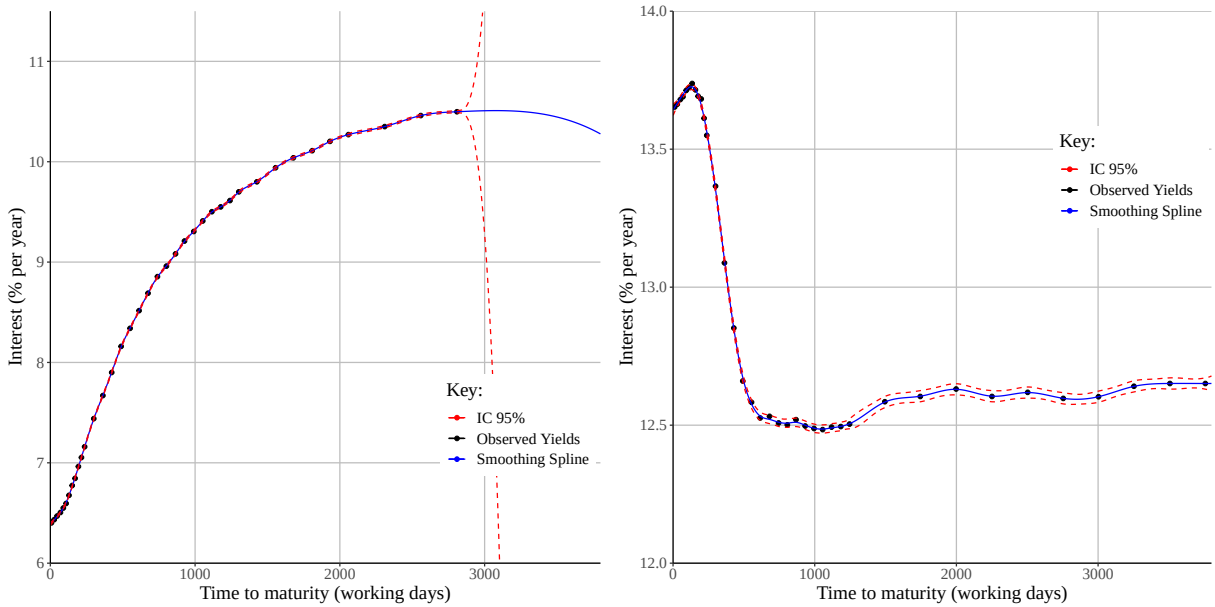
(a) Normal Yield Curve - $\lambda = 0$ **(b)** Inverted Yield Curve - $\lambda = 0$

Source: Elaborated by the author.

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 (λ). 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

[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.

Figure 2.5: Smoothing splines - λ by GCV

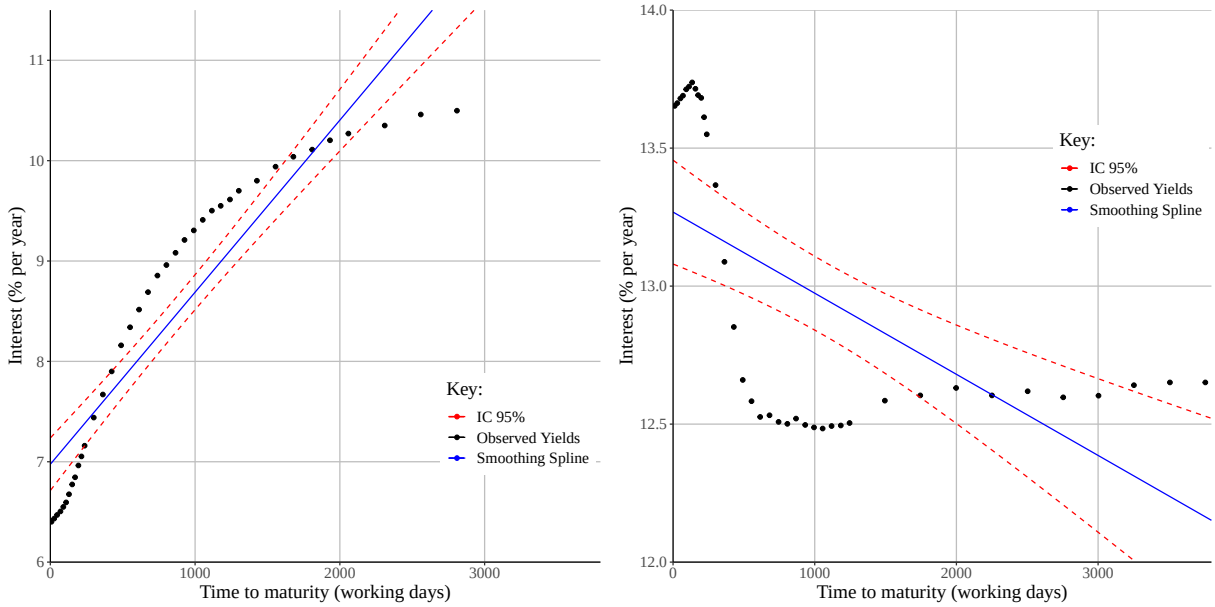


(a) Normal Yield Curve - λ by GCV

(b) Inverted Yield Curve - λ by GCV

Source: Elaborated by the author.

Figure 2.6: Smoothing splines - $\lambda = 100$



(a) Normal Yield Curve - $\lambda = 100$

(b) Inverted Yield Curve - $\lambda = 100$

Source: Elaborated by the author.

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 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 (2.12); notice that maturity is measured in years in this case:

$$RSS(f,\lambda) = \sum_{i=1}^{M} \left[y_i - f(m_i)\right]^2 + \lambda(m) \int \left[f''(m)\right]^2 dt \tag{2.11}$$

$$\lambda(m) = \begin{cases} 0.1 & 0 \leq m \leq 1 \\ 100 & 1 \leq m \leq 10 \\ 100.000 & 0 \leq 10 \leq m \end{cases} \cdot \tag{2.12}$$

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, accordingly with Formula (2.13). The parameters in this function (L, S, and μ) were initially estimated from the maximisation of out-of-sample goodness-of-fit. However, the authors obtained multiple maxima from different parameter values, so they chose those that produced the highest level of smoothness:

$$ln[\lambda(m)] = L - (L - S) \times e^{\left(\frac{-m}{\mu}\right)}. \tag{2.13}$$

2.2.2 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 section delves into 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.

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 $M = m$:

$$g(m) = E(Y|M = m) = \int y f_{Y|M=m}(y)dy = \frac{\int y f_{M,Y}(m, y)dy}{f_M(m)}. \tag{2.14}$$

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,

[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).

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

$$\hat{f_M}(m) = \frac{\hat{F}_n(m+h) - \hat{F}_n(m-h)}{2h} = \frac{\sum_{i=1}^{N} \mathbb{1}_{[m-h,m+h]}(m_i)}{2hn} \tag{2.15}$$

$$= \frac{1}{nh}\sum_{i=1}^{N} \underbrace{\frac{\mathbb{1}_{[m-h,m+h]}(m_i)}{2}}_{\text{(I)}}. \tag{2.16}$$

The term (I) in equation (2.16) 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):

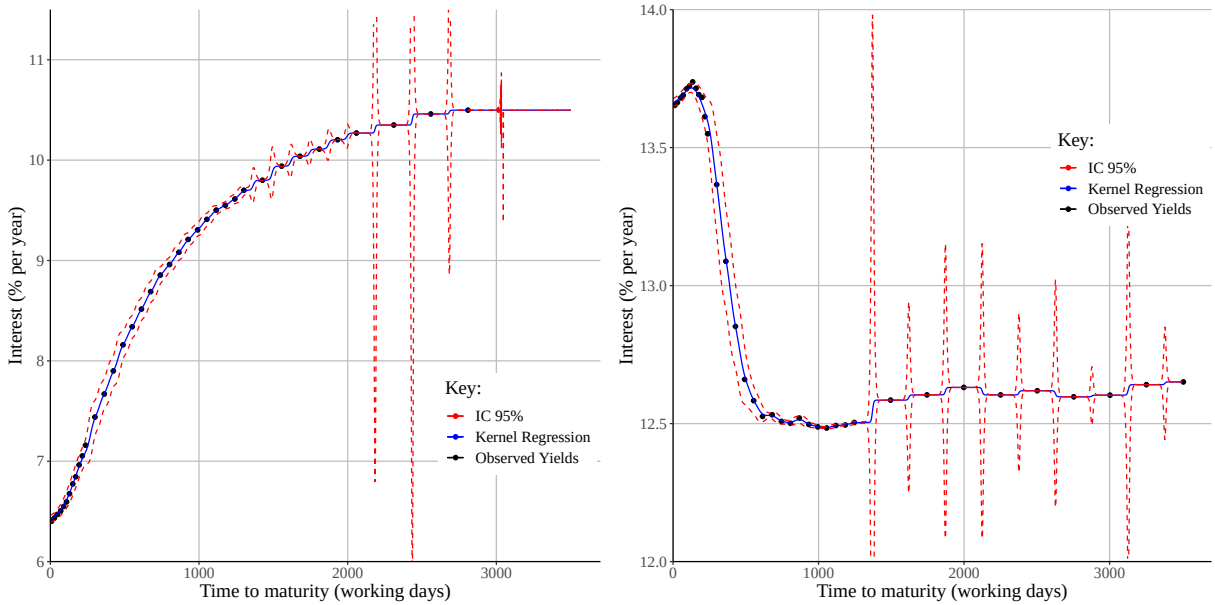
$$\hat{g}(m) = \frac{\int y f_{X,Y}(x,y)dy}{f_X(x)} = \frac{\sum_{i=1}^{N} y_i K_\lambda\left(\frac{m_i - m}{\lambda}\right)}{\sum_{i=1}^{N} K_\lambda\left(\frac{m_i - m}{\lambda}\right)}. \tag{2.17}$$

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. These choices play a crucial role in determining the accuracy and characteristics of the Nadaraya-Watson estimator's results.

Figure (2.7) illustrates yield curve estimates obtained through the Nadaraya-Watson estimator, utilising a Gaussian kernel and bandwidth selection via least-squares cross-validation, as described by Hayfield and Racine (2008). However, the chosen bandwidth selection method apparently overfits the data and results in unstable confidence bands at specific data points for both curves. In contrast, Figure (2.8) presents the estimates while fixing the bandwidth at 0.3, which successfully eliminates the issue of overfitting and produces stable confidence bands.

Naturally, this model extrapolation maintains its stability, as it replicates the last local average for the additional maturities.

Figure 2.7: Nadaraya and Watson Kernel Regression - CV bandwidth = 0.0809



(a) Normal Yield Curve **(b)** Inverted Yield Curve

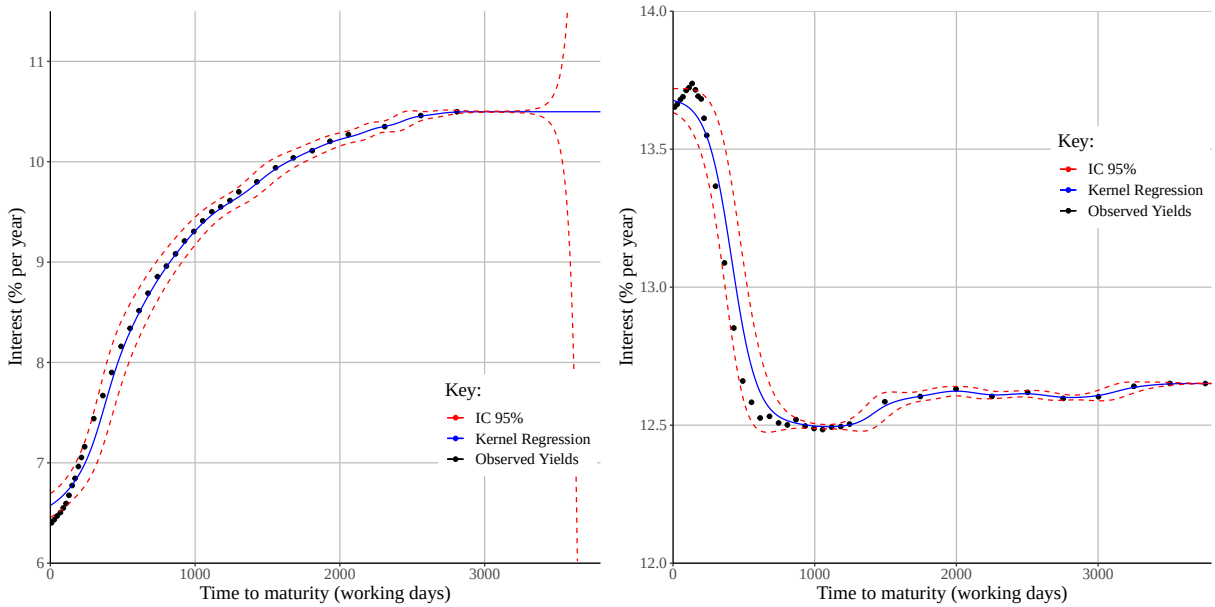
Source: Elaborated by the author.

Local regression (Loess)

As evident in the short-end of the yield curves in Figure (2.8), 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, as exemplified in Figure (2.8b) for maturities ranging from 0 to 1000. 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.

Figure 2.8: Nadaraya and Watson Kernel Regression - bandwidth = 0.3



(a) Normal Yield Curve **(b)** Inverted Yield Curve

Source: Elaborated by the author.

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.

Algorithm - Local regression

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

  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.

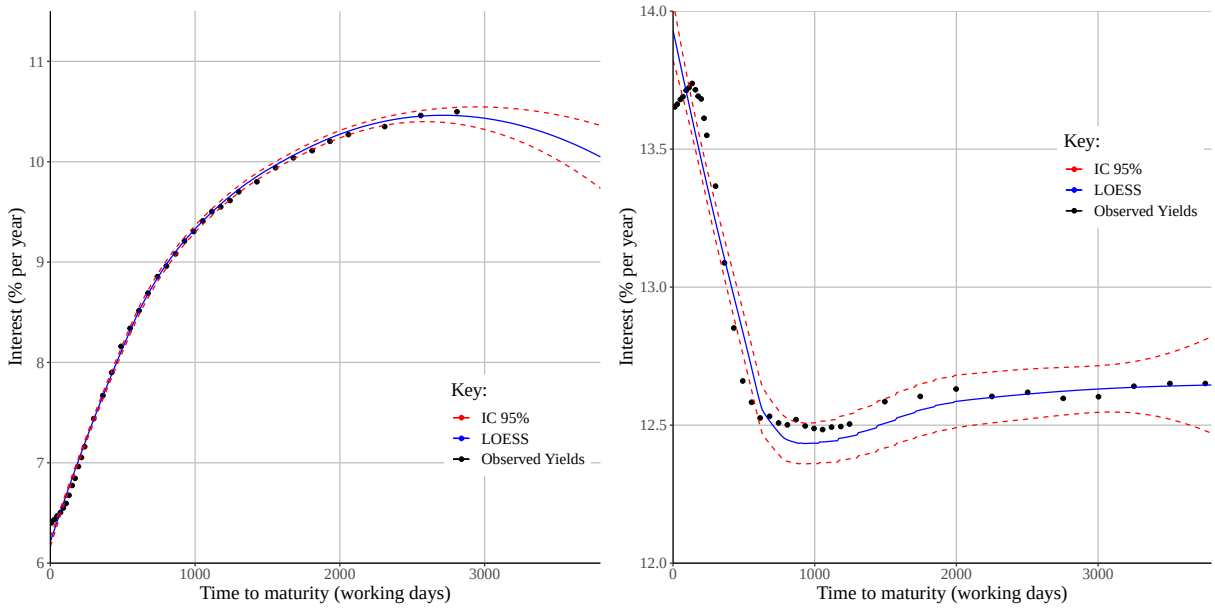
3. Estimate a weighted least squares regression considering the weights assigned in step 2:

$$\min_{\beta_0,\beta_1} \sum_{j=1}^{N} K(m_i, m_j)(y_j - \beta_0 - \beta_1 x_j)^2$$

4. Get the estimate: $\hat{m}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

Yield curve estimates obtained with Loess are presented in Figure (2.9) setting a span of 20% and using tricubic weighting[7]. What is noteworthy in these estimates is the downward trend in the extrapolated values, as in Figure (2.9a), which is not appropriate given the underlying economic theory.

Figure 2.9: Loess



(a) Normal Yield Curve **(b)** Inverted Yield Curve

Source: Elaborated by the author.

[7] $K(v) = (1 - (distance_v - distance_max)^3)^3$.

2.2.3 Nelson-Siegel family models

The polynomial regression discussed in Section 2.2.1 seeks to model the entire maturity domain using a single function. However, as mentioned earlier, 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.

In addition to its role as an estimation technique, this approach serves as an initial step in modelling yield curve dynamics, as proposed by Diebold and Li (2006). When examining a single working day with a cross-section of yields and maturities, the models presented in this section require parameter estimation. In the context of dynamic modelling, estimates of function parameters over a sequence of working days can be modelled as a time series, enabling yield curve forecasting. However, this exercise is beyond the scope of this chapter.

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 (2.18), which comprises a constant term and a Laguerre function, which is characterised by a polynomial multiplied by an exponential decay term, as follows[8]:

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1-e^{-m/\lambda_t}}{m/\lambda_t}\right) + \beta_{2,t}\left(\frac{1-e^{-m/\lambda_t}}{m/\lambda_t} - e^{-m/\lambda_t}\right) + \epsilon_{m,t}. \tag{2.18}$$

In Equation (2.18), 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:

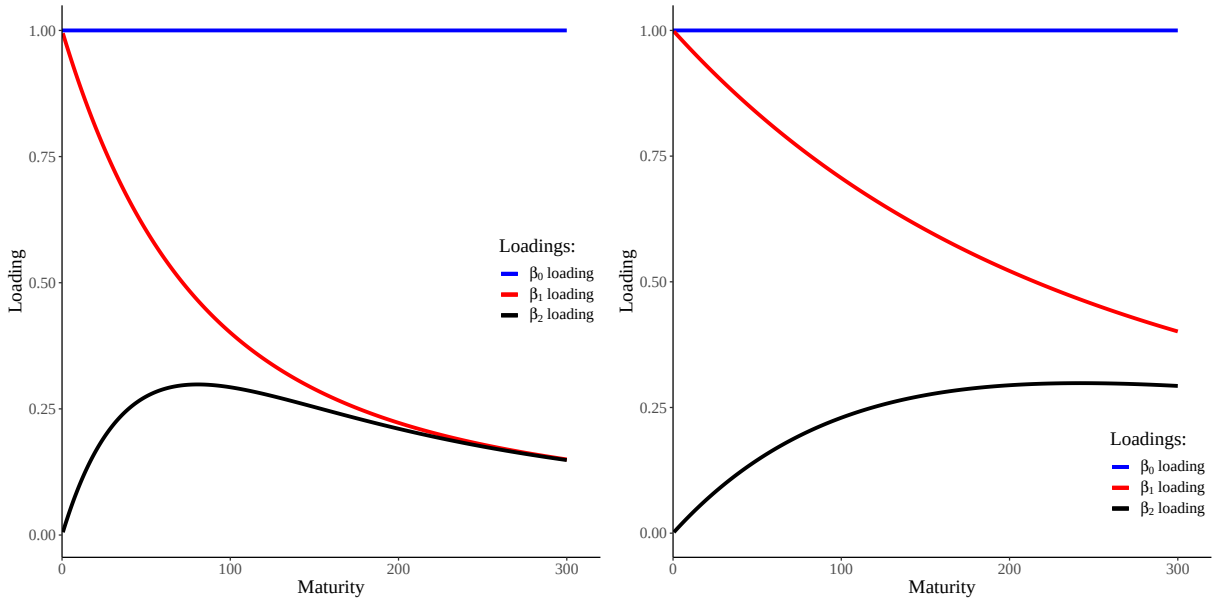
- Factor $\beta_{0,t}$ is associated with the long-run interest rate and controls the overall level of the function.
- Factor $\beta_{1,t}$ is linked to the short-run interest rate and governs the function's inclination. The factor loading makes it most influential when $m = 0$, decaying monotonically 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 $\beta_{2,t}$). Figure 2.10 presents how the factor loadings evolve with the maturity for two distinct λ_t levels. It's

[8]The specification presented in Equation (2.18) refers to the yield-to-maturity TSIR representation; it is possible to obtain equivalent specifications for the other forms following the relations presented in Appendix A.

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).

Figure 2.10: Evolution of Nelson-Siegel factor loadings over time



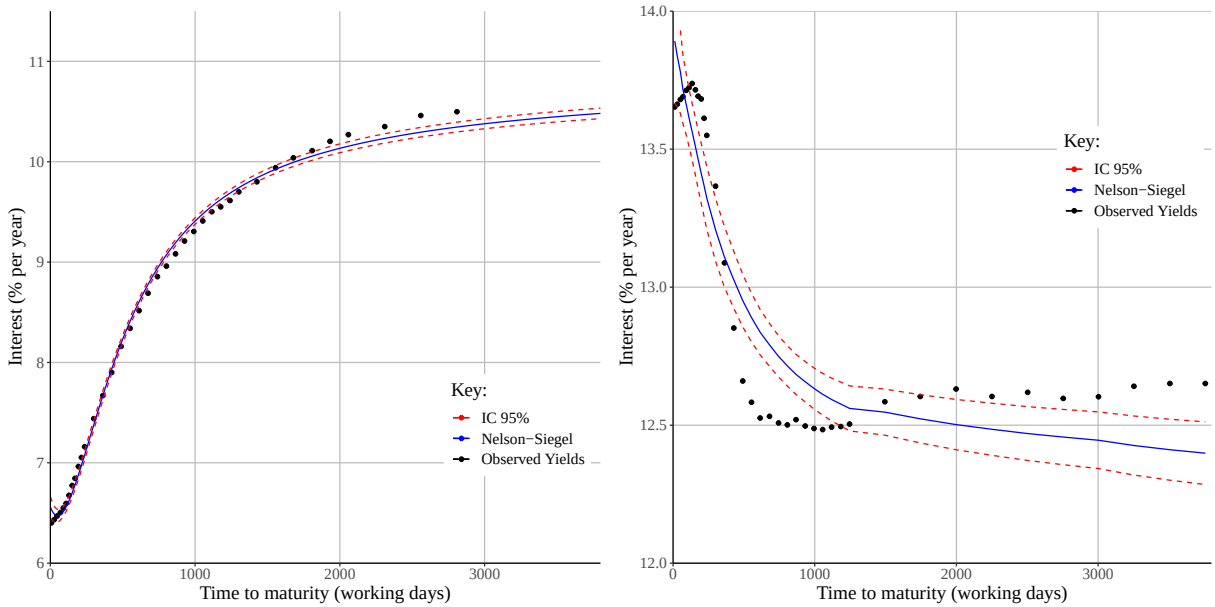
(a) $\lambda_t = 45$ **(b)** $\lambda_t = 130$

Source: Elaborated by the author.

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 (2.18) results in over-parametrisation without substantial gains in model fit.

Treating λ_t as an unknown parameter alongside the factors (βs) implies that Equation (2.18) 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. Two examples of the Nelson-Siegel model fitted through ordinary least squares, with $\lambda_t = 150$, is displayed in Figure 2.11. Contrastingly, two estimates for the same days using nonlinear least squares are presented in Figure 2.12.

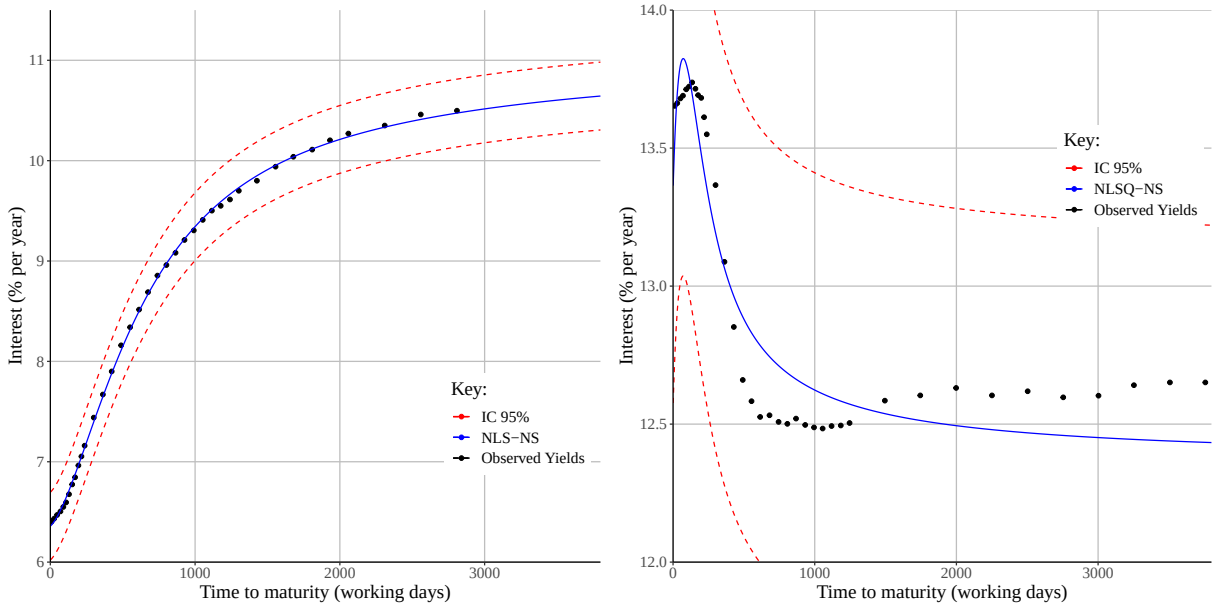
Figure 2.11: Nelson-Siegel ordinary least squares fit



(a) Normal Yield Curve **(b)** Inverted Yield Curve

Source: Elaborated by the author.

Figure 2.12: Nelson-Siegel nonlinear least squares fit



(a) Normal Yield Curve **(b)** Inverted Yield Curve

Source: Elaborated by the author.

In their initial 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 delve into 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. The impact of this phenomenon is demonstrated in Table 2.1 using synthetic data, revealing alterations in the correlation between factor loadings with varying λ_t values. Subsequently, Table 2.2 holds the λ_t value constant, allowing an assessment of how the correlation changes with distinct maturity vectors.

Table 2.1: Factor loadings correlation: dependence on λ.

	$\lambda = 1.37$			$\lambda = 3$			$\lambda = 10$		
	FL1	FL2	FL3	FL1	FL2	FL3	FL1	FL2	FL3
FL1	1			1			1		
FL2		1	0.084		1	-0.524		1	-0.966
FL3		0.084	1		-0.524	1		-0.966	1

Factor loadings generated with different λ values for the same maturity vector, λ set in a scale for yearly data. Synthetic maturity vector used for illustration.

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 coef-

Table 2.2: Factor loadings correlation: dependence on the maturity vector.

	m_1			m_2		
	FL1	FL2	FL3	FL1	FL2	FL3
FL1	1			1		
FL2		1	0.084		1	0.985
FL3		0.084	1		0.985	1

Factor loadings generated with $\lambda = 1.37$ set in a scale for yearly data.
Synthetic maturity vector used for illustration.

ficient estimates exhibiting an inverted signal concerning the underlying theory, coupled with inaccuracies in their magnitude. Table 2.3 provides a glimpse into these effects, illustrating two estimates for the Nelson-Siegel specification. Both estimates use identical yield-maturity data, and the variations in λ_t results in substantial fluctuations in the coefficient estimates ($\beta_{1,t}$, $\beta_{2,t}$).

Table 2.3: Regression Estimation Sensitivity to Correlated Factor Loadings

	Dependent variable:	
	$y(m)$	
	(1)	(2)
FL1	9.090***	9.243***
	(0.204)	(0.192)
FL2	4,625.388***	4.186
	(1,073.282)	(2.624)
FL3	−4,721.613***	−22.596***
	(1,090.307)	(4.893)
Observations	38	38
R^2	0.986	0.989
Adjusted R^2	0.985	0.988
Residual Std. Error (df = 35)	1.059	0.957
F Statistic (df = 3; 35)	822.277***	1,008.807***
Note:		*p<0.1; **p<0.05; ***p<0.01

Factor loadings generated with different λ values (1.37, 10) for the same data points.
Data from 22 October 2018, the normal curve considered in this section.

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 2.18 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 2.18. 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 2.18 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.

The Bliss model

According to Bliss (1996, p. 11-12), the over-parametrisation identified by Nelson and Siegel in equation (2.18) was attributed 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 (2.19),

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}}\right) + \beta_{2,t}\left(\frac{1-e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}} - e^{-m/\lambda_{2,t}}\right) + \epsilon_{m,t}. \tag{2.19}$$

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.[9]

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. Therefore, he proposed a specification – equation (2.20) – with a second medium-term factor, with a specific decay parameter. His intent was to model a second "hump" on the yield curve,

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

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}}\right) + \beta_{2,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}} - e^{-m/\lambda_{1,t}}\right) + \beta_{3,t}\left(\frac{1-e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}} - e^{-m/\lambda_{2,t}}\right) + \epsilon_{m,t}. \quad (2.20)$$

The Five-factors model

A further specification in the Nelson-Siegel family considered in this chapter 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.[10] 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 and Ferreira (2013); Rezende (2011) 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:

$$y_t(m) = \beta_{0,t} + \beta_{1,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}}\right) + \beta_{2,t}\left(\frac{1-e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}}\right) + \beta_{3,t}\left(\frac{1-e^{-m/\lambda_{1,t}}}{m/\lambda_{1,t}} - e^{-m/\lambda_{1,t}}\right) + \beta_{4,t}\left(\frac{1-e^{-m/\lambda_{2,t}}}{m/\lambda_{2,t}} - e^{-m/\lambda_{2,t}}\right) + \epsilon_{m,t}. \quad (2.21)$$

[10] 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. These models are not discussed any further in this chapter.

2.3 Model comparison criteria

To compare estimates from the models presented in the previous section, it is imperative to define metrics that enable the evaluation of various dimensions by which a yield curve estimate can be assessed. Relevant literature[11] 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):

$$MSE = \frac{\sum_{i=m_1}^{M}(y_i - \hat{y}_i)^2}{M} \qquad MAE = \frac{\sum_{i=m_1}^{M}|y_i - \hat{y}_i|}{M}. \tag{2.22}$$

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.

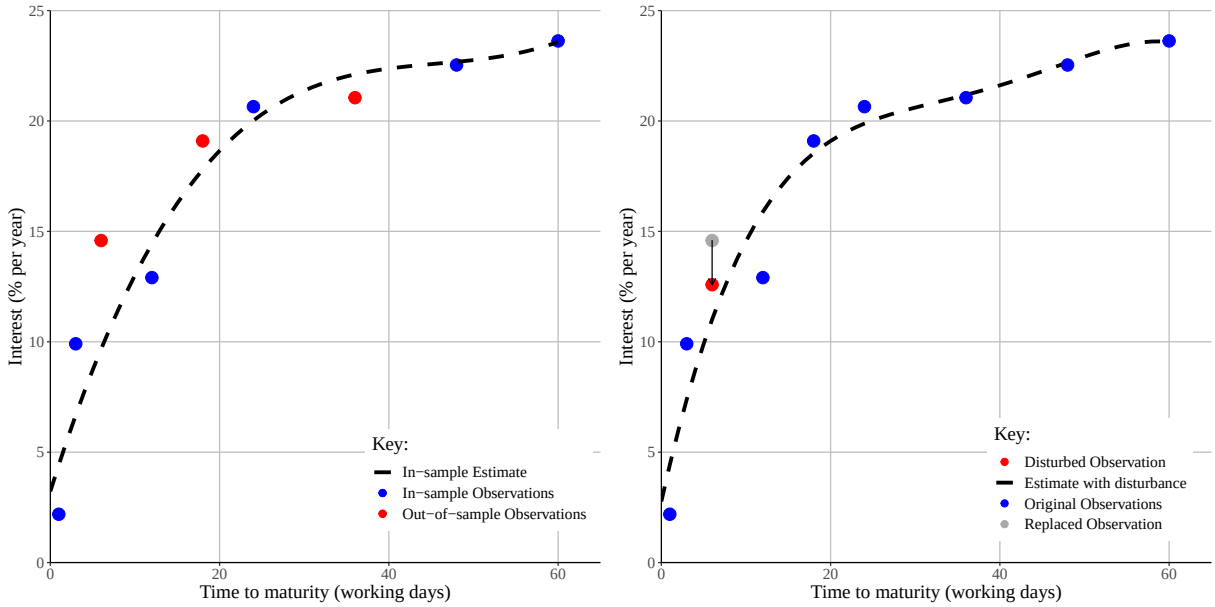
To overcome these difficulties, we consider a **training set** (or in-sample set) and a **validation set** (or out-of-sample set) for each day in the sample. A typical working day has around 38 data points; we classify these points into three groups according to maturity: maturities up to one year, maturities above one year and below three years, and maturities over three years.

[11] 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.

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 and its divisions. Figure 2.13a 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 2.13: 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[12] model would exhibit a significant increase in these measures, indicating that the disturbance has affected the estimates across multiple maturities. Figure 2.13b illustrates this process with synthetic data. The third 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

$$R(y_t) = \int_{m_1}^{M} [D^2 y(s)]^2 \, ds. \tag{2.23}$$

The roughness measure in equation (2.23) 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.[13] Seppälä and Viertiö (1996, p. 21) argue that since we have more information at the

[12] "*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)

[13] 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

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 (2.23) 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

$$R_2(y_t) = \int_{m_1}^{M} [s \times D^2 y(s)]^2 \, ds. \tag{2.24}$$

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,

$$R_3(y_t) = \frac{1}{M - m_1} \int_{m_1}^{M} [D^2 y(s)]^2 \, ds. \tag{2.25}$$

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

2.3.1 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 re-

in this text.

quires 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:

$$S = \frac{12D}{K(K+1)} \sum_{k=1}^{K} \left(\bar{R}_k - \frac{K+1}{2} \right)^2 \underset{D\to\infty}{\sim} \chi^2_{K-1}. \tag{2.26}$$

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

$$|\bar{R}_{k_1} - \bar{R}_{k_2}| \geq r_{\alpha,K,D}, \tag{2.27}$$

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 (2.27) 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.[14]

To implement the multiple comparison tests, we follow Koning et al. (2005) once again, who used plots to present the results compactly. In subsection 2.5, 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

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

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.

2.4 Data and estimation procedures

The data used in this chapter comprises of rates of One-day Interbank Deposit Futures ("DI1"). 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.

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 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 (see Appendix A) 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.

2.5 Results: comparing estimation techniques

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 dimensions explained in section 2.3 and test whether the performances of the models are statistically different.

2.5.1 Goodness-of-fit analysis

In this section, we analyse the goodness-of-fit of all models. Starting with descriptive statistics, Table 2.4 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 2.5 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. This is remarkable since we do not know of previous applications of Loess to yield curve estimation.

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.

The key takeaways from Tables 2.4 and 2.5 are:

- The Smoothing Spline has the best fit in-sample but is outperformed by the Loess out-of-sample.
- The Kernel regression specifications fit out-of-sample data poorly.
- The Nelson-Siegel family has a similar fit, but the Nelson-Siegel specification estimated by OLS has a poor fit in the short range.
- The Polynomial regression fit is not much worse than that of the Nelson-Siegel family.

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 2.6 confirm that the models have different performances. Then, we implement the **multiple comparison procedure** on Figures 2.14 and 2.15.

From Figure 2.14a, 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 2.14b, 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 2.14c and 2.14d), 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 2.15a and 2.15b). 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 2.15c and 2.15d).

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 2.16 and 2.17 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 2.18 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.

Table 2.4: Assessing Model Goodness-of-fit: Mean Absolute Error (MAE) 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	0.118840	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

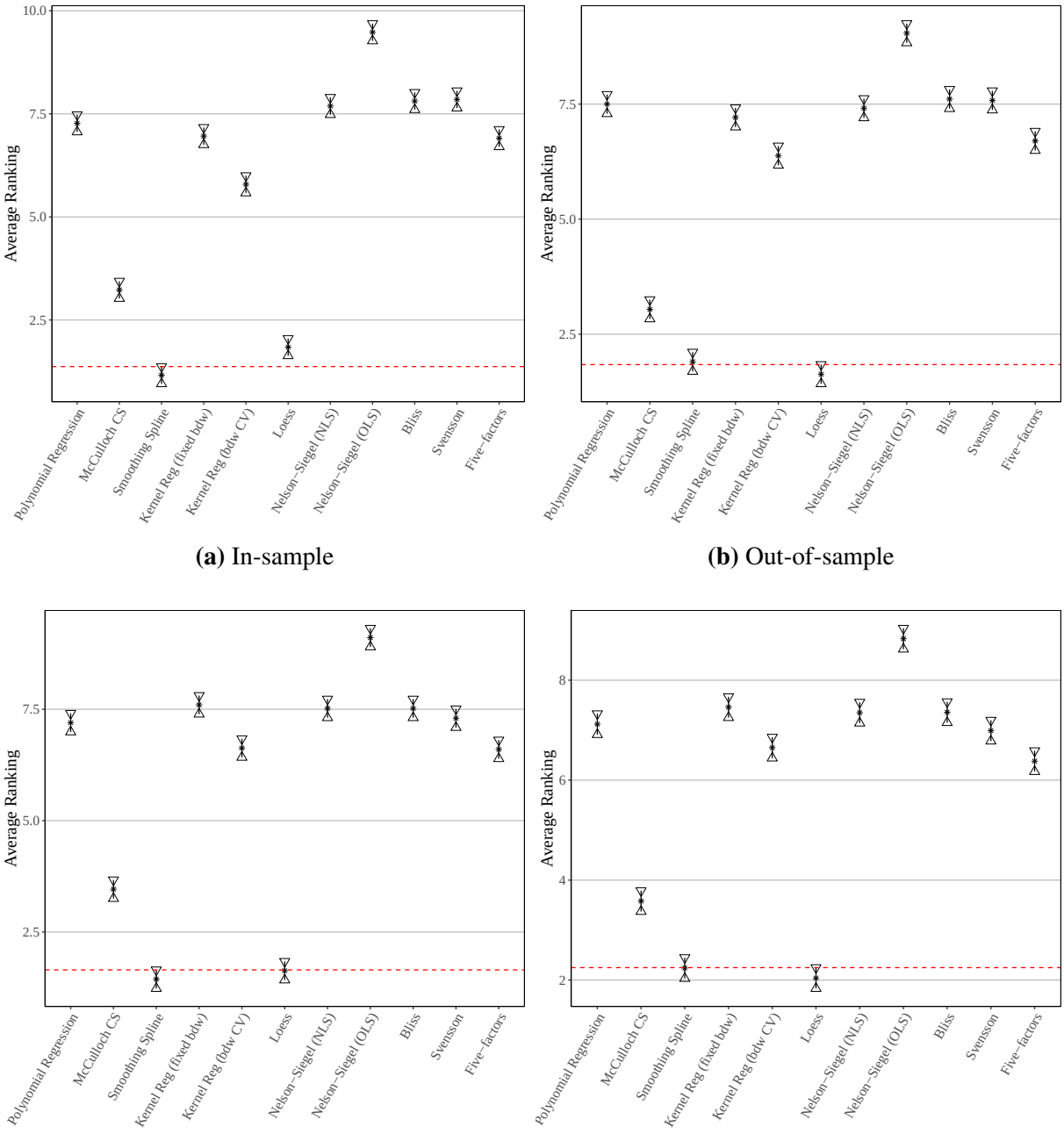
Table 2.5: Assessing Model Goodness-of-fit: Mean Square Error (MSE) in- and out-of-sample

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 2.6: Friedman rank sum test - Models ranked by MAE

Avg Ranking	Full Sample	In-sample	Out-of-sample	In-sample (Short Range)	Out-of-sample (Short Range)	In-sample (Medium Range)	Out-of-sample (Medium Range)	In-sample (Long Range)	Out-of-sample (Long Range)	Pert. Sample
Friedman S	9382.039	8777.956	7791.033	8005.149	6243.955	8023.111	6075.483	10330.403	6829.723	5306.144
df	10	10	10	10	10	10	10	10	10	10
p-value	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16

Figure 2.14: Multiple Comparison Procedure for Goodness-of-Fit - Models Ranked by MAE (Overall and Short Range)



(a) In-sample

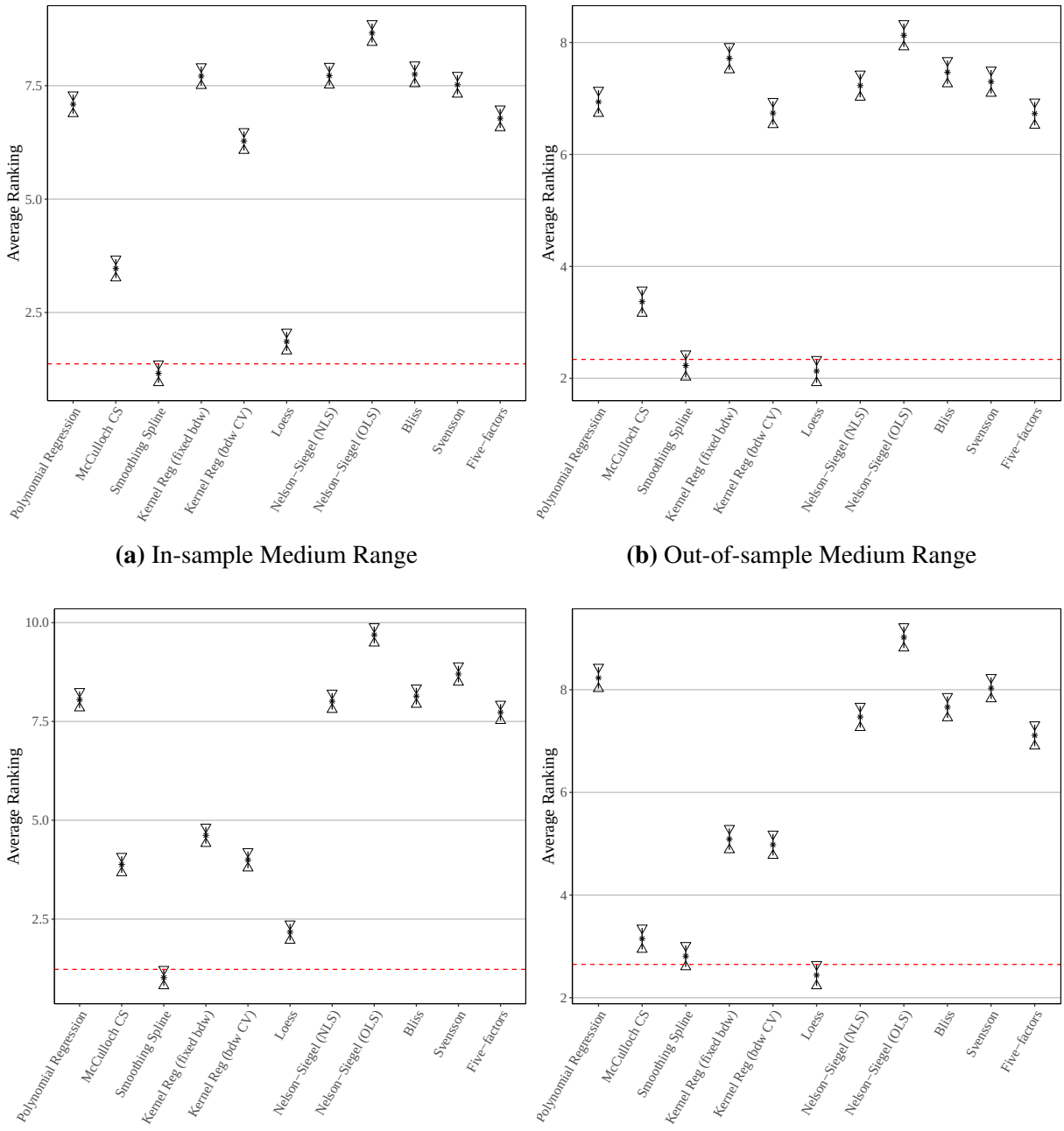
(b) Out-of-sample

(c) In-sample Short Range

(d) Out-of-Sample Short Range

Source: Elaborated by the author.

Figure 2.15: Multiple Comparison Procedure for Goodness-of-Fit - Models Ranked by MAE (Medium Range and Long Range)



(a) In-sample Medium Range

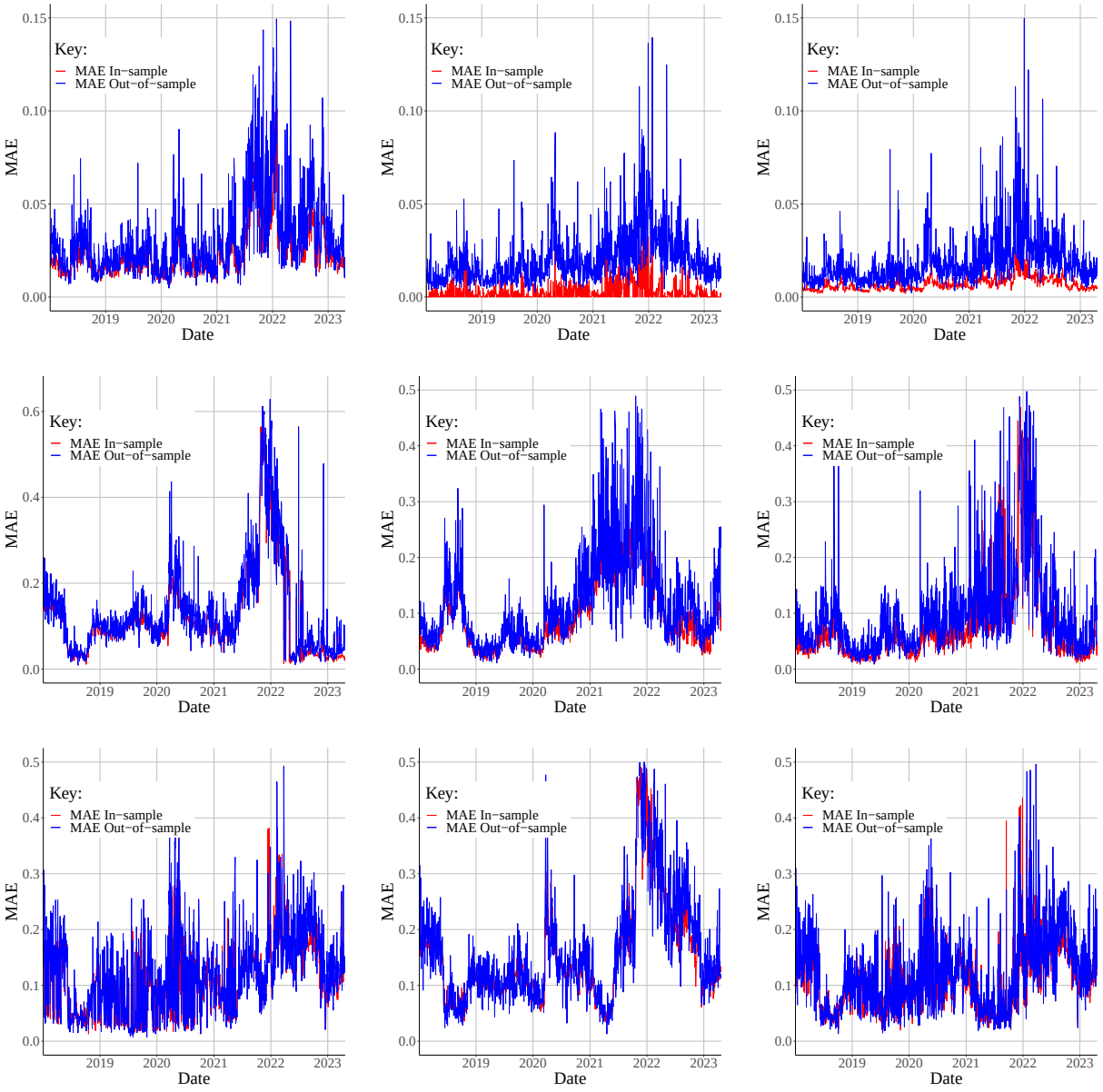
(b) Out-of-sample Medium Range

(c) In-sample Long Range

(d) Out-of-Sample Long Range

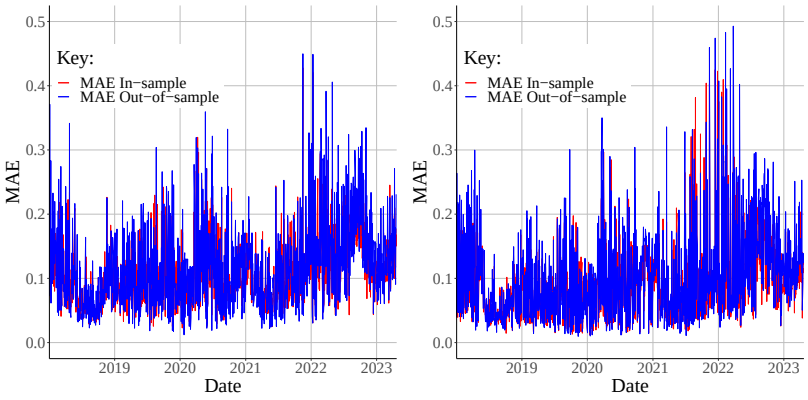
Source: Elaborated by the author.

Figure 2.16: 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.

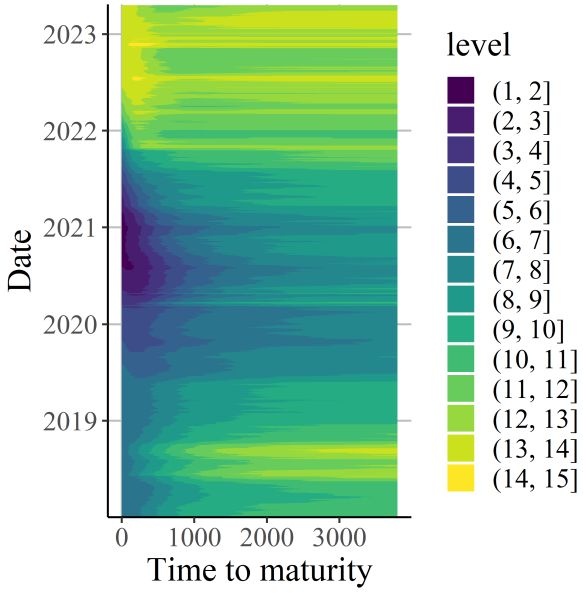
Figure 2.17: In-sample and Out-of-sample Accuracy Across Sample Days (2 of 2)



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

Source: Elaborated by the author.

Figure 2.18: The Yield Curve Evolution: January 2018 - April 2023



Source: Elaborated by the author.

2.5.2 Robustness analysis

This section analyses how the model fit changes when one outlier is introduced to the sample. Table 2.7 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.

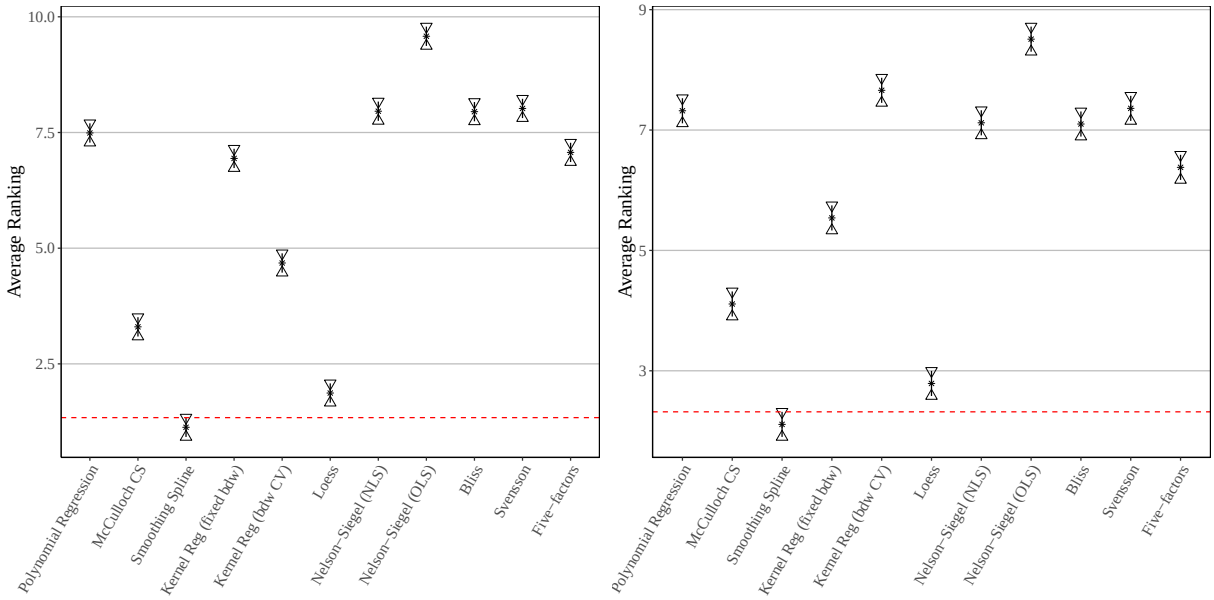
Table 2.7: Assessing Model Robustness: MAE and MSE in Original and Perturbed Sets

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

The Friedman test statistic in Table 2.6 indicates significant differences among the models in both the original sample and the sample with disturbance. The multiple comparison test in Figure 2.19 reveals that, except for the Kernel regression specifications, the average rankings of the models remain unchanged with the introduction of the outlier, and the smoothing spline maintains the best-fit rank in both settings.

For the robustness analysis, we also depict the evolution of MAE over the sample days, considering both estimates in Figures 2.20 and 2.21. The graphical analysis highlights that the disturbance led to an upward shift in the average MAE across the days. This effect is particularly noticeable in the spline models and Loess. However, for the Nelson-Siegel family, this effect is not as pronounced.

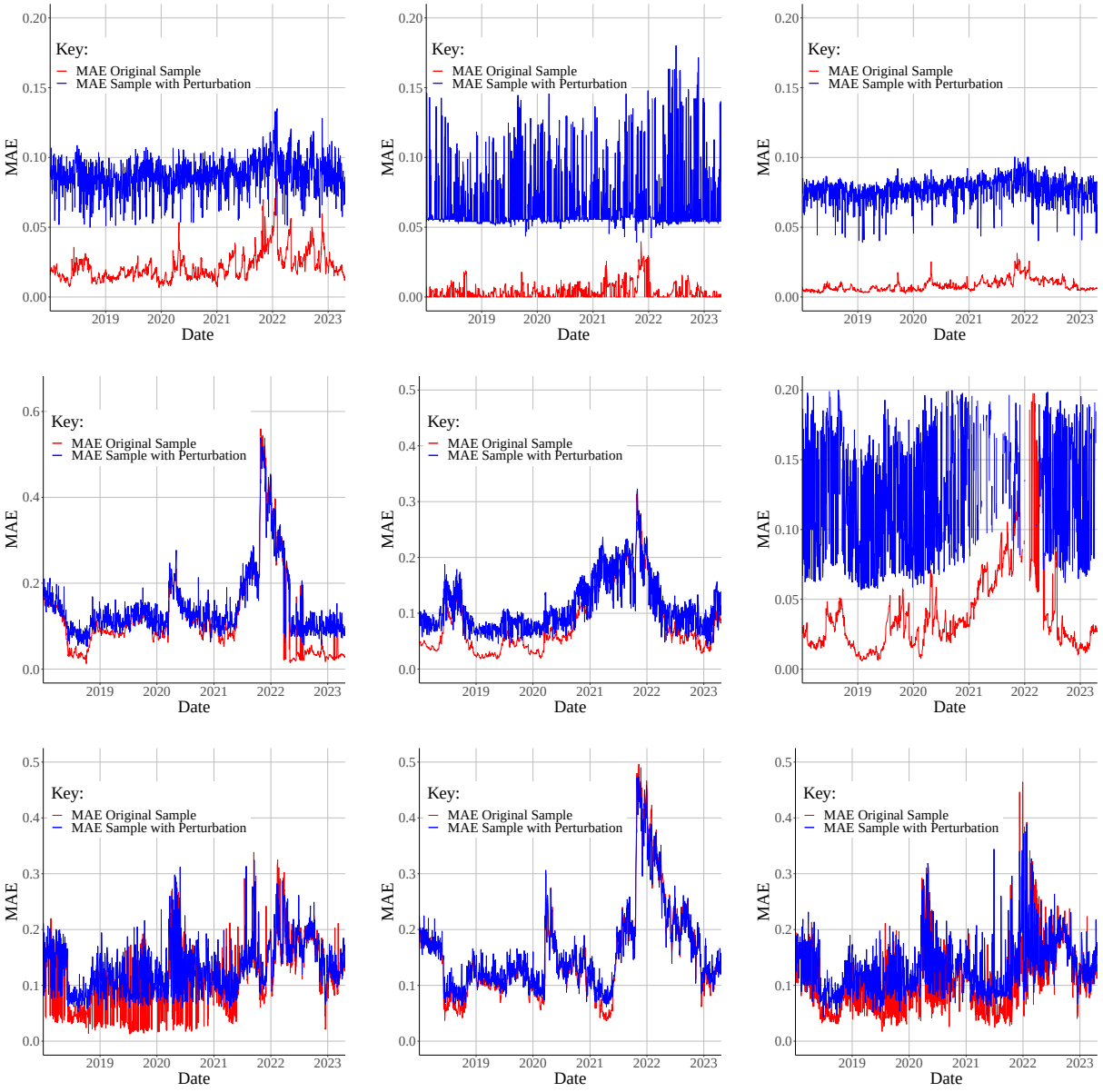
Figure 2.19: Multiple Comparison Procedure for Robustness - Models Ranked by MAE



(a) Original Sample **(b)** Sample with disturbance

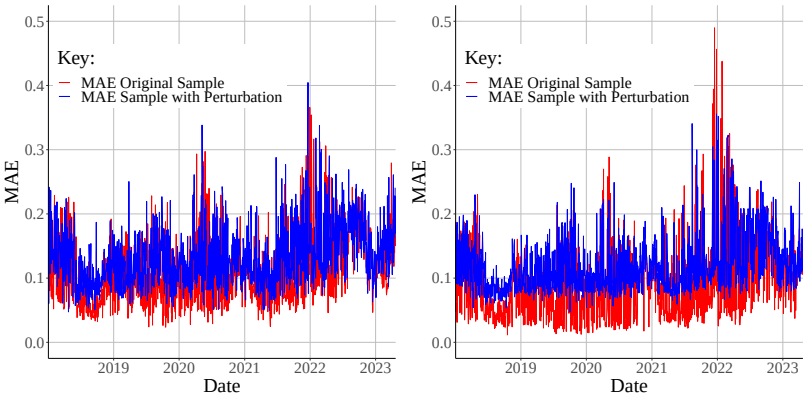
Source: Elaborated by the author.

Figure 2.20: Robustness 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.

Figure 2.21: Robustness Across Sample Days (2 of 2)



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

Source: Elaborated by the author.

2.5.3 Smoothness analysis

This section analyses how smooth the models' estimates are according to the three roughness metrics presented previously. Table 2.8 provides the average metrics considering all days in the sample.

Considering R, Polynomial, Svensson, and Loess specifications present the lower roughness. 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 maturity, aligning more with the example in Figure 2.4 than that in Figure 2.5. In other words, a high R_2 for the smoothing spline indicates poor extrapolation after the last observation.

Table 2.8: 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

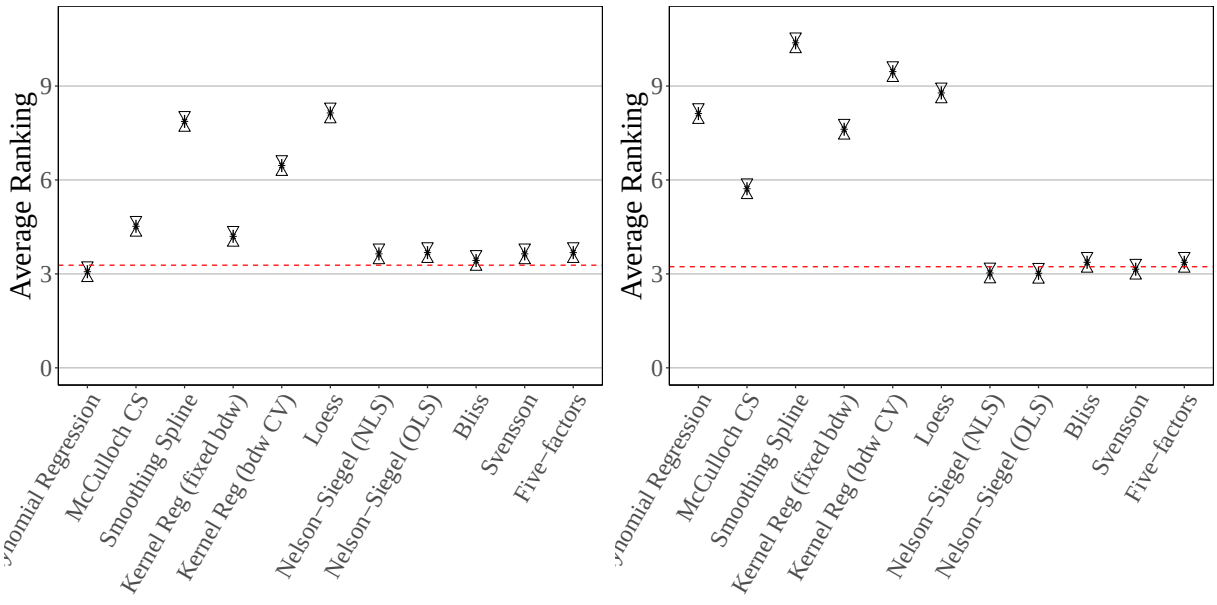
Ranking the models according the smoothness metrics and applying the Friedman tests in-

dicates that models' performance in terms of roughness are indeed different.[15]

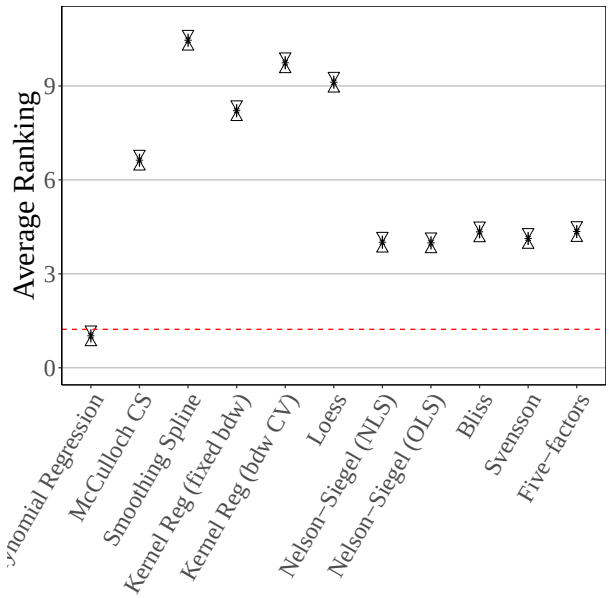
Figure 2.22 shows the **multiple comparison procedure**. Estimates within the Nelson-Siegel family are not statistically different in any setting. Considering the rank by R, the Nelson-Siegel family and the Polynomial regression systematically have the lowest roughness in the sample. Considering R_2, the Nelson-Siegel family is unchallenged, while the rankings given by R_3 show the Polynomial regression with the fewest rough estimates.

[15]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 2.22: Multiple Comparison Procedure for Smoothness - Models Ranked by R, $R2$, $R3$



(a) R **(b)** R_2



(c) R_3

Source: Elaborated by the author.

2.6 Conclusion

This chapter provides a comprehensive evaluation of empirical models of the Term Structure of Interest Rates, commonly known as the Yield Curve. The Yield Curve represents the relationship between time to maturity and the level of interest rates, playing a pivotal role in economic analysis. However, the Yield Curve is not fully observable on a daily basis. Each day, one only observes interest levels at specific maturities corresponding to the bonds or derivatives contracts traded that day. The models analysed in this chapter aim to interpolate the observed points and extrapolate over the last observed maturity to estimate the complete Yield Curve.

Literature usually considers nonparametric or spline models in addition to the parsimonious function models, derived from Nelson and Siegel (1987)'s seminal work. Even though the chapter surveyed more spline models, only the two primary tools of the trade were considered for the comparison of the estimates: McCulloch's Natural Cubic Spline and Smoothing Splines. On the other hand, the chapter analysed four specifications from the Nelson-Siegel family and two 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.

Chapter 3

Forecasting the Yield Curve: an implementation of the Dynamic Nelson-Siegel model

3.1 Introduction

Comprehending the yield curve dynamics holds significant importance for policy-makers and private investors. On the one hand, having insights into the expected behaviour of interest rates in the coming months or years aids governments in optimising their debt structures. Additionally, more precise forecasts of the yield curve empower policy-makers to enhance the execution of monetary policy by anticipating market reactions to overnight interest rates and other macroeconomic variables. From the private sector perspective, these forecasts enable institutional investors to manage risk in their portfolios effectively, thereby enhancing their long-term investment strategies.

This chapter delves into an examination of the modelling and forecasting capabilities of the Dynamic Nelson-Siegel model, as proposed by Diebold and Li (2006) and further developed in Diebold et al. (2006), in comparison to baseline models (random walk and Holt-Winters)

for forecasting the term structure of interest rates in Brazil. The landscape of models for term structure is diverse, with various theoretical and empirical approaches. While theoretical models might face challenges accurately fitting real-world data, empirical models, although effective in fitting data, often lack a solid economic foundation, making their interpretation for policy purposes complex. Additionally, only a few models estimate and forecast the term structure. The Diebold-Li specification, along with its variations, stands out as an advantageous option, exhibiting both good forecasting performance in various scenarios and interpretable parameters.

In this chapter, we utilise Brazilian data obtained from interest rate futures contracts ("DI1") spanning from January 2018 to April 2023 to estimate and forecast the term structure of interest rates. The raw data is subjected to interpolation using natural cubic splines, ensuring that all models operate with interest rates at predefined maturities as their input. The chosen time frame encompasses various shapes of the Yield Curve, providing a comprehensive basis for evaluating the forecasting performance of the models.

This chapter is organised into four sections, each serving a distinct purpose alongside this introduction. Section 3.2 provides an overview of the Diebold-Li specification and the methodologies used for its estimation. Additionally, it reviews previous studies that have employed this framework for forecasting the Brazilian Yield Curve. Section 3.3 provides detailed information about the data used in the forecasting exercises and outlines the criteria for comparing the forecasts. Section 3.4 reveals the out-of-sample performance of the models, and Section 3.5 concludes this chapter.

3.2 Methods for forecasting yield curve

The model introduced by Nelson and Siegel (1987) gave rise to a new category of yield curve estimation techniques known as *parsimonious* parametric models, characterised by their few parameters requiring estimation. Not only do these models fit interest rate data well, but their parameters are also easily interpretable, contributing to their popularity among central banks

and money market agents. Three model parameters are associated with the short- (β_1), medium- (β_2), and long-term (β_0) components of the term structure, the fourth parameter (λ) is a decay factor that influences the maturity at which the medium-term factor reaches its maximum:

$$y(m) = \beta_0 + \beta_1 \left(\frac{1 - e^{-m/\lambda}}{m/\lambda} \right) + \beta_2 \left(\frac{1 - e^{-m/\lambda}}{m/\lambda} - e^{-m/\lambda} \right) + \varepsilon_m. \tag{3.1}$$

When faced with a cross-section of yield-maturity pairs, the model can be estimated, and yield estimates for unobserved maturities can be obtained using the model. However, the presence of the parameter λ in equation (3.1) forces the utilisation of nonlinear least squares to estimate all four parameters. Nelson and Siegel suggested employing a grid of λ, then estimating the other parameters using ordinary least squares (OLS) and selecting the appropriate value for λ based on the sum of squared residuals (SSR).

Later, Diebold and Li (2006) proposed a dynamic version of the original model,

$$y_t(m) = \beta_{0,t} + \beta_{1,t} \left(\frac{1 - e^{-m/\lambda}}{m/\lambda} \right) + \beta_{2,t} \left(\frac{1 - e^{-m/\lambda}}{m/\lambda} - e^{-m/\lambda} \right) + \varepsilon_{m,t}. \tag{3.2}$$

In this instance, they set the value for λ constant across all days in their sample, taking into account the average maturity at which the curvature factor reaches its maximum.[1] By considering a series of Nelson-Siegel parameter estimates over a sequence of days or months, they could be treated as a set of time series:

$$\begin{cases} \{\hat{\beta}_{0,t}\} = \hat{\beta}_{0,1}, \hat{\beta}_{0,2}, \hat{\beta}_{0,2}, \dots \\ \{\hat{\beta}_{1,t}\} = \hat{\beta}_{1,1}, \hat{\beta}_{1,2}, \hat{\beta}_{1,2}, \dots \\ \{\hat{\beta}_{2,t}\} = \hat{\beta}_{2,1}, \hat{\beta}_{2,2}, \hat{\beta}_{2,2}, \dots \end{cases} \tag{3.3}$$

Therefore, one could model and forecast the parameters, resulting in a forecast of the yield

[1] Diebold and Li fixed the decay parameter ($\lambda_t = \lambda; \forall t$, see the Appendix A) and estimated the others using OLS, see section 2.2.3 for the Nelson-Siegel estimation details.

curve. Furthermore, Diebold and Li provided an alternative interpretation for these parameters as latent factors, aligning with Litterman and Scheinkman (1991). The latter authors proposed summarising the yield variability over different maturities into a few factors that govern the yield curve's shape. In this context, the Nelson-Siegel parameters are associated with the yield curve's level (β_0), steepness (β_1), and curvature (β_2). To illustrate this association, Diebold and Li (2006, p. 341 and 350) suggest comparing Nelson-Siegel factor estimates with empirical (or data-based) measures of level, inclination, and curvature.[2]

Originally, Diebold and Li (2006) introduced a two-step estimation procedure. In the first step, equation (3.2) is estimated by OLS (i.e. with a fixed λ) for a sequence of days or months. Subsequently, the resulting series (3.3) are modelled as following three first-order autoregressive processes or a first-order a vector autoregressive process. Then the yield curve forecasts are produced using the formula

$$\hat{y}_{t+h|t}(m) = \hat{\beta}_{0,t+h|t} + \hat{\beta}_{1,t+h|t}\left(\frac{1-e^{-m/\lambda}}{m/\lambda}\right) + \hat{\beta}_{2,t+h|t}\left(\frac{1-e^{-m/\lambda}}{m/\lambda_t} - e^{-m/\lambda}\right), \tag{3.4}$$

where:

$$\hat{\beta}_{i,t+h|t} = \hat{a}_i + \hat{\gamma}_i\hat{\beta}_{i,t}; \qquad i = 0,1,2 \qquad \text{for the AR(1) processes, or}$$

$$\hat{\boldsymbol{\beta}}_{t+h|t} = \hat{\alpha} + \hat{\boldsymbol{\gamma}}\hat{\boldsymbol{\beta}}_t \qquad \text{for the VAR(1) process.}$$

In the context of AR(1) modelling, one can proceed in two ways to obtain h-steps ahead forecasts. One approach involves considering the original data frequency and producing multi-step forecasts from 1 to h. Alternatively, as suggested by Diebold and Li (2006), estimates $\hat{a}_i$ and $\hat{\gamma}_i$ can be obtained by regressing $\hat{\beta}_{i,t}$ on an intercept and $\hat{\beta}_{i,t-h}$, aiming to minimise the forecast

[2]Specifically, one should observe: $\beta_{1,t} = y_t(\infty)$ and $-\beta_{2,t} = y_t(\infty) - y_t(0)$, where ∞ correspond to the longest maturity observed on a given working day. Further, $\beta_{3,t} \approx 2y_t(18) - y_t(1) - y_t(48)$. This comparison is made on figure 3.3, contrasting Nelson-Siegel factors estimated daily by OLS and the data-based measures given by: Level $= y_t(120)$, Slope $= -(y_t(120) - y_t(3))$, and Curvature $= \frac{1}{0.3} \times 2y_t(24) - y_t(3) - y_t(120)$; with maturity measured in months.

root mean squared error (RMSE). Once the parameters are forecasted, yields for the whole maturity range are calculated using the formula (3.4). Then, one assesses the forecasting quality by comparing these calculated yields with out-of-sample yield observations (in $t + h$) rather than comparing parameter forecasts in period t with parameter estimates in period $t + h$.

While the Diebold and Li (2006) forecasts outperformed competing models in many maturity ranges, the two-step procedure had some inconsistencies. Specifically, the series of factors modelled as autoregressive processes are non-stationary. Furthermore, the determination of λ is somewhat *ad hoc*. To address the inconsistencies, Diebold et al. (2006) proposed a one-step estimation procedure. Starting with the VAR representation of the Nelson-Siegel parameter series and writing equation (3.2) in matrix form, one gets:

$$\underbrace{\begin{pmatrix} y_t(m_1) \\ y_t(m_2) \\ \vdots \\ y_t(m_M) \end{pmatrix}}_{\boldsymbol{y}(m)} = \underbrace{\begin{pmatrix} 1 & \frac{1-e^{-m_1/\lambda_t}}{m_1/\lambda_t} & \frac{1-e^{-m_1/\lambda_t}}{m_1/\lambda_t} - e^{-m_1/\lambda_t} \\ 1 & \frac{1-e^{-m_2/\lambda_t}}{m_2/\lambda_t} & \frac{1-e^{-m_2/\lambda_t}}{m_2/\lambda_t} - e^{-m_2/\lambda_t} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-m_1/\lambda_t}}{m_M/\lambda_t} & \frac{1-e^{-m_M/\lambda_t}}{m_M/\lambda_t} - e^{-m_M/\lambda_t} \end{pmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{pmatrix} \beta_{0,t} \\ \beta_{1,t} \\ \beta_{2,t} \end{pmatrix}}_{\boldsymbol{\beta}_t} + \underbrace{\begin{pmatrix} \varepsilon_t(m_1) \\ \varepsilon_t(m_2) \\ \vdots \\ \varepsilon_t(m_M) \end{pmatrix}}_{\boldsymbol{\varepsilon}_t} \tag{3.5}$$

$$\underbrace{\begin{pmatrix} \beta_{0,t} - \mu_0 \\ \beta_{1,t} - \mu_1 \\ \beta_{2,t} - \mu_2 \end{pmatrix}}_{\boldsymbol{\beta}_t - \boldsymbol{\mu}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} \beta_{0,t-1} - \mu_0 \\ \beta_{0,t-1} - \mu_1 \\ \beta_{0,t-1} - \mu_2 \end{pmatrix}}_{\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}} + \underbrace{\begin{pmatrix} \eta_{0,t} \\ \eta_{1,t} \\ \eta_{2,t} \end{pmatrix}}_{\boldsymbol{\eta}_t}. \tag{3.6}$$

In a more concise form using vector-matrix notation, the system of equations modelling the parameters' evolution and the yield curve estimation in the state-space representation can be expressed as:

$$\mathbf{y}_t = \boldsymbol{\Lambda}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t \qquad \text{Measurement equation} \tag{3.7}$$

$$(\boldsymbol{\beta}_t - \boldsymbol{\mu}) = \boldsymbol{A}(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t \qquad \text{State equation.} \tag{3.8}$$

Contrasting with models studied in chapter 2, besides estimating the term structure each day with equation (3.7) the dynamic Nelson-Siegel specification also models the term structure evolution over the days with equation (3.8). To estimate this system using the **Kalman filter**, Diebold et al. (2006) assumes that the error components in the measurement and transition equations are orthogonal to each other and that the parameters in the first period ($t = 0$) are orthogonal to the error term in the correspondent equation:

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim WN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H & 0 \\ 0 & Q \end{pmatrix} \right]; \qquad \begin{cases} E(\boldsymbol{\beta}_0 \boldsymbol{\varepsilon}_t') = 0 \\ E(\boldsymbol{\beta}_0 \boldsymbol{\eta}_t') = 0 \end{cases} \tag{3.9}$$

Further assumptions are that the errors at different maturities are uncorrelated (**H** is diagonal), while errors affecting level, slope, and curvature may be correlated (**Q** is nondiagonal).[3] Remarkably, this approach enables the daily estimation of the four parameters in the measurement equation (level, slope, curvature, and the decay parameter, λ_t). Furthermore, other potentially relevant variables in the yield curve evolution, such as inflation and activity level, may be included in the state equation.

3.2.1 The estimation process using the Kalman Filter

The Kalman filter is a recursive procedure to obtain the optimal estimator of the state vector ($\boldsymbol{\beta}_t$) at time t, using information available up to time t, (Harvey, 1989, p. 104). The system of equations (3.7) and (3.8) combined with the assumptions (3.9) constitute a linear Gaussian state-space model.

Following Koopman et al. (2010), the state vector ($\boldsymbol{\beta}_t$) can be estimated using past and current values of the observed yields ($\mathbf{y}_1$, $\mathbf{y}_2$, ..., $\mathbf{y}_t$). Let $\mathbf{b}_{t|s}$ represent the Minimum Mean Square Linear Estimator (MMSLE) of $\boldsymbol{\beta}_t$[4], and $\mathbf{B}_{t,s}$ denote the Mean Square Error (MSE) matrix

[3] "[Then] *application of the Kalman filter then delivers maximum-likelihood estimates and optimal filtered and smoothed estimates of the underlying factors.*", Diebold et al. (2006, p. 313).

[4] As presented by Harvey (1989, p. 110-111): $\mathbf{b}_t = E_t(\boldsymbol{\beta}_t) = E(\boldsymbol{\beta}_t|\mathbf{y}_t)$

for $s = t - 1$. With the knowledge of the values of $\mathbf{b}_{t|t-1}$ and $\mathbf{B}_{t,t-1}$, the Kalman filter updates these values by incorporating the new yield observations $\mathbf{y}_t$ during the **filtering step**:

$$\mathbf{b}_{t|t} = \mathbf{b}_{t|t-1} + \mathbf{B}_{t|t-1}\mathbf{\Lambda}'\mathbf{F}_t^{-1}\mathbf{v}_t, \tag{3.10}$$

$$\mathbf{B}_{t|t} = \mathbf{B}_{t|t-1} - \mathbf{B}_{t|t-1}\mathbf{\Lambda}'\mathbf{F}_t^{-1}\mathbf{\Lambda}\mathbf{B}_{t|t-1} \tag{3.11}$$

Where:

- $\mathbf{\Lambda}$ is the loading factors matrix, as in Formula (3.5);
- $\mathbf{v}_t = \mathbf{y}_t - \Lambda\mathbf{B}_{t|t-1}$, is the prediction error one-step ahead; and
- $\mathbf{F}_t = \Lambda\mathbf{B}_{t|t-1}\Lambda' + \mathbf{Q}$, is the prediction error variance matrix.

After updating the estimates of $\boldsymbol{\beta}_t$, the **prediction step** of the filter computes the MMSLE of the state vector for the subsequent period,

$$\mathbf{b}_{t+1|t} = (\mathbf{I} - \mathbf{A})\boldsymbol{\mu} + \mathbf{A}\mathbf{b}_{t|t}, \tag{3.12}$$

$$\mathbf{B}_{t+1|t} = \mathbf{A}\mathbf{B}_{t|t}\mathbf{A}' + \mathbf{Q}. \tag{3.13}$$

Where $\mathbf{A}$ is the coefficients matrix from the state equation.

Considered jointly, the updating equations (3.10 and 3.10) the prediction equations (3.12 and 3.13) constitute the Kalman filter. Let's consolidate all unknown parameters, including the coefficients in $\mathbf{A}$, the λ_t in $\mathbf{\Lambda}$, $\boldsymbol{\mu}$, and the unknown matrices ($\mathbf{H}$ and $\mathbf{Q}$), into a parameter vector ($\mathbf{\Psi}$). Considering the distributional assumptions in (3.9) one can write the likelihood function for the sample, following Harvey (1989, p. 125-126):

$$L(\mathbf{y}, \mathbf{\Psi}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}) \tag{3.14}$$

$$\ell(\mathbf{\Psi}) = -\frac{NT}{2}ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T} ln|\mathbf{F}_t| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t. \tag{3.15}$$

Then, starting with initial values ($\boldsymbol{\beta}_1 = \boldsymbol{\mu}$, and $\mathbf{B}_{1,0}$), the unknown parameters are estimated recursively over the sample of days $t = 1, \ldots, T$. For each day t, the parameter estimates are obtained by maximising the log-likelihood function in relation to $\boldsymbol{\Psi}$.

3.2.2 Related studies for the Brazilian Yield Curve

In their comparative analysis, Vicente and Tabak (2008) evaluated the performance of the dynamic Nelson-Siegel model using Brazilian data, employing the two-step procedure over a 12-month horizon. They compared two Nelson-Siegel specifications – one with a fixed decay and the other with the decay parameter estimated alongside others by NLS – with two affine term structure models (ATMS) specifications and random walk ("*no change*") forecasts. Through the assessment of mean squared errors in out-of-sample forecasts, they concluded that the Diebold and Li (2006) model exhibited superior performance, particularly for longer time spans, as determined by the Diebold-Mariano test (Diebold and Mariano, 1995).

Similarly, Cajueiro et al. (2009) conducted a comparison of Diebold-Li forecast performance (using the two-steps procedure) with the Functional Signal Plus Noise (FSPN) model developed by Bowsher and Meeks (2006) and random walk forecasts. Utilising the Diebold-Mariano test, they concluded that while the FSPN model provided better forecasts up to three months, it was outperformed by the Diebold-Li specification in longer maturities.

Employing Brazilian data, Almeida et al. (2009) applied a dynamic version of the Svensson (1994) specification[5] using the two-step estimation procedure proposed by Diebold and Li (2006). They conducted a forecast comparison between the dynamic Nelson-Siegel and the dynamic Svensson models, considering AR(1) and VAR(1) processes, along with random walk forecasts. Both the Nelson-Siegel and Svensson specifications utilised fixed decay parameters, determined as the value minimising in-sample root mean square error (RMSE). This work found that the dynamic Svensson model produced better estimates in three forecast horizons (one day,

[5] An extension of the baseline Nelson-Siegel model, incorporating a second curvature (medium-term) factor. Refer to section 2.2.3 for details on the Svensson model.

one month, and three months ahead) according to the Diebold-Mariano test.

The study conducted by Caldeira et al. (2010) estimated the Diebold-Li specification in the state-space form using the Kalman filter with Brazilian data. They fixed the decay parameter for the two-step estimating procedure, ensuring that the curvature factor reached its maximum in maturities between 13 and 18 months. In comparing the two-step and one-step forecasts three and six months ahead, they found that the latter estimation produced superior out-of-sample forecasts for RMSE, as indicated by the Diebold-Mariano test.

The study conducted by Carvalho and Moura (2014) compared the performance of dynamic versions of the Nelson-Siegel and Svensson models, estimated using the two-step procedure with AR(1) and VAR(1) structures. The comparison included a two-factor model (a Nelson-Siegel specification without the curvature term) and a random walk forecast. The findings indicated that the dynamic Nelson-Siegel and Svensson models were outperformed, as measured by out-of-sample RMSE, by the random walk and two-factor models on many horizons and maturities considered.

The study by Caldeira et al. (2016) implemented an arbitrage-free dynamic Nelson-Siegel (AFNS) model to estimate and forecast the Brazilian yield curve. Both the conventional Diebold-Li and the AFNS specifications were estimated using the Kalman filter. The estimation process involved a rolling window of 2 years (500 working days), and the models generated forecasts for various time horizons: one week, one month, three months, and six months ahead. The performance of these models was compared against forecasts generated by the random walk, AR(1), and VAR(1) models. The results showed that while the Diebold-Li and AFNS models performed better at longer horizons, they were outperformed by other models in shorter horizons.

Finally, the study conducted by Vieira et al. (2017) explored the impact of additional macroeconomic information on forecasting the Brazilian yield curve. Rather than incorporating a few extra factors, as done by Diebold et al. (2006) in the state equation, they condensed a substantial number of macroeconomic variables using principal components. Consequently, the

state equation adopted the form of a factor-augmented VAR. They registered an improvement in forecast accuracy over other models, mainly at short maturities.

3.3 Data

Similar to Chapter 2, this chapter relies on data related to rates of One-day Interbank Deposit Futures ("DI1") from 2 January 2018 to 22 April 2023. This dataset covers various yield curve shapes, offering a comprehensive foundation for assessing forecasting performance. For each working day, the last *Price Report* made available by B3 in XML files was downloaded, using the R package `RSelenium` resulting in a sample of 1313 working days. Typically, a day has 37 or 38 records, each corresponding to a different DI1 maturities. 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.

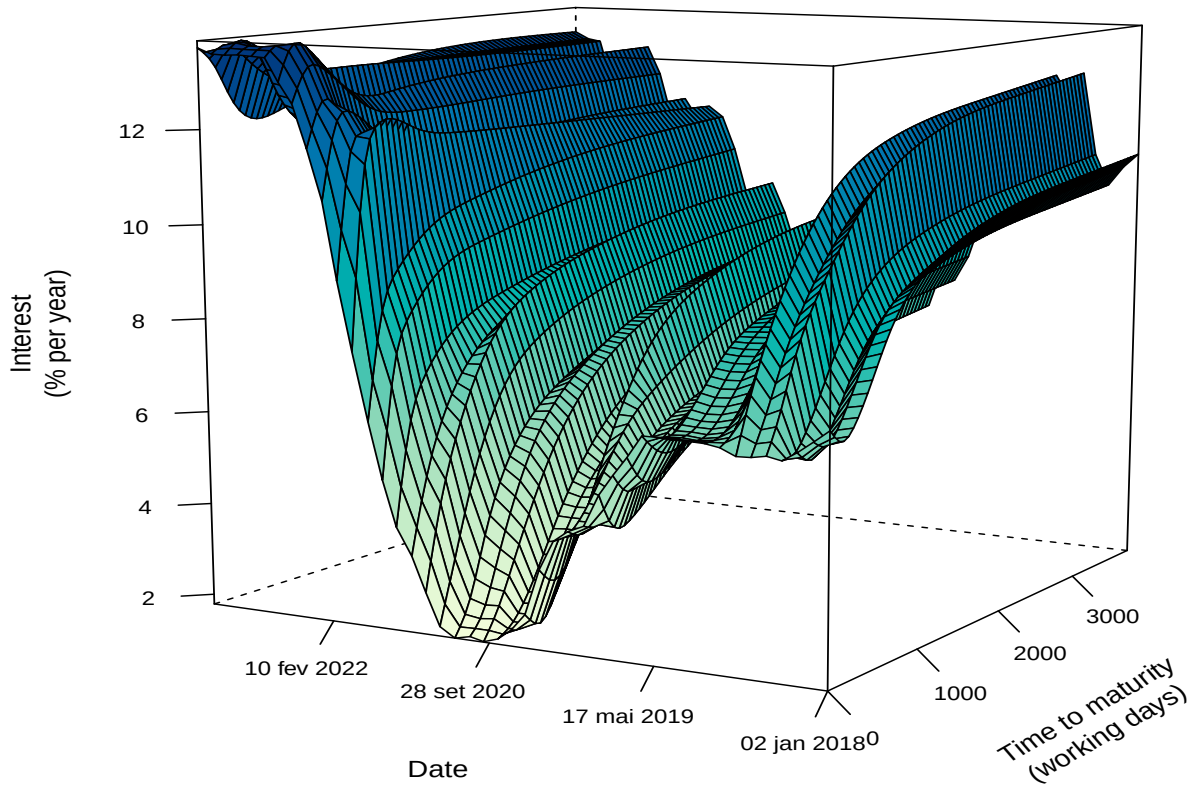
Due to the daily variation in available maturities, this chapter adopts the complete term structure estimated by McCulloch's Natural Cubic Spline (refer to Section 2.2.1). Daily yields were extracted for maturities of 3, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. While not mandatory, this is a standard procedure in literature.[6]

Figure 3.1 presents how yield curve evolved over our sample period: starting with a normal (positively inclined) curve that changed its level, becoming a humped curve in 2022, and finally an inverted curve in 2023.

Table 3.1 present the descriptive statistics for the selected maturities, as well as for the data-based measures of level, slope and curvature. Considering the column with average yields at different maturities, one can observe that the "average yield curve" is increasing and concave, characteristic of a normal yield curve. Furthermore, the standard error column aligns with the

[6] Similar datasets have been utilised in previous studies such as Vicente and Tabak (2008), Cajueiro et al. (2009), Almeida et al. (2009), Caldeira et al. (2010), and Caldeira et al. (2016). Standardising maturities by estimating the yield curve in advance is a common practice in literature, even though it is not a prerequisite for the primary estimation procedure addressed in this chapter. For instance, Diebold and Li (2006) used linear interpolation to standardise the maturities.

Figure 3.1: Brazilian Term Structure, January 2018 - April 2023



Note: McCulloch Natural Cubic Splines Estimates.
Source: Elaborated by the author.

stylised fact of greater variability in the short end compared to the long end. The autocorrelation estimates at lags of 1, 6, and 12 months are presented; the high values of these statistics indicate a high persistence of interest rates at these maturities. On the other hand, the partial autocorrelations are relatively low. Assessing the descriptive statistics of the empirical factors, one can observe that the averages are theoretically coherent (level > 0 and level + slope > 0, for instance), and the factors exhibit a high degree of persistence.

Table 3.1: Descriptive statistics for the Yield Curves - January 2018 - April 2023

Maturity (months)	$\bar{X}$	s.e.	min	max	$\hat{\rho}(1)$	$\hat{\rho}(6)$	$\hat{\rho}(12)$	$pacf(1)$	$pacf(6)$
3	7.0730	3.9083	1.8729	13.8884	0.9713	0.6868	0.1529	-0.0078	-0.0040
6	7.2976	3.9606	1.8881	14.2306	0.9724	0.7008	0.1879	-0.0017	-0.0006
9	7.4763	3.9101	1.9953	14.4259	0.9724	0.7074	0.2078	0.0046	0.0028
12	7.6207	3.7643	2.1858	14.4899	0.9717	0.7101	0.2188	0.0075	0.0041
18	7.8878	3.4102	2.6826	14.4019	0.9684	0.7097	0.2297	0.0115	-0.0002
24	8.1505	3.0902	3.2339	14.1820	0.9635	0.7052	0.2325	0.0166	-0.0005
30	8.3998	2.8270	3.7675	13.9414	0.9577	0.6976	0.2298	0.0186	0.0042
36	8.6252	2.6378	4.2093	13.7848	0.9518	0.6882	0.2243	0.0175	0.0086
48	8.9954	2.4270	4.9328	13.7466	0.9433	0.6697	0.2119	0.0128	0.0152
60	9.2749	2.3093	5.4841	13.7802	0.9375	0.6536	0.2002	0.0097	0.0203
72	9.4839	2.2241	5.9054	13.7709	0.9329	0.6404	0.1899	0.0100	0.0235
84	9.6460	2.1576	6.2385	13.7374	0.9289	0.6294	0.1815	0.0116	0.0254
96	9.7744	2.1040	6.4328	13.7076	0.9253	0.6199	0.1744	0.0124	0.0266
108	9.8756	2.0609	6.5485	13.7433	0.9224	0.6117	0.1687	0.0129	0.0275
120	9.9555	2.0262	6.6521	13.7686	0.9200	0.6049	0.1641	0.0129	0.0280
Level	9.9555	2.0262	6.6521	13.7686	0.9200	0.6049	0.1641	0.0129	0.0280
Slope	-2.8825	2.4533	-6.5232	2.2814	0.9395	0.5532	-0.0873	0.0248	-0.0009
Curvature	-0.7274	1.3021	-3.6680	3.8600	0.8426	0.4195	-0.0904	0.0270	-0.0007

3.4 Empirical Results

This section employs the one-step procedure (the Kalman filter) to estimate the dynamic Nelson-Siegel model. This approach aims to evaluate the latent factor interpretation proposed by Diebold and Li. Additionally, we conduct a comparative analysis of the forecasting accuracy between the dynamic Nelson-Siegel model estimated via the two-step procedure, where coefficients are modelled as either an AR(1) process or a VAR(1) process. This comparison involves assessing the model against two baseline forecasting techniques: the random walk forecast, where predictions correspond to the last observation in the training set and an exponential smoothing technique known as Holt-Winters.

3.4.1 Latent factors interpretation

Diebold and Li (2006) introduced an alternative interpretation for the coefficients within the Nelson-Siegel formula. Initially, Nelson and Siegel had linked these coefficients to interest com-

ponents of various maturities (long, medium, and short-term components). However, Diebold and Li suggested an alternative perspective, positing that these coefficients could be construed as latent factors intrinsic to yield curve characteristics: level, slope, and curvature. Each factor is influenced by its respective factor loading (denoted as matrix Λ in Formula (3.5)). The assessment of the latent factor interpretation is presented in Table 3.2 as correlation coefficients and graphically in Figure 3.3.

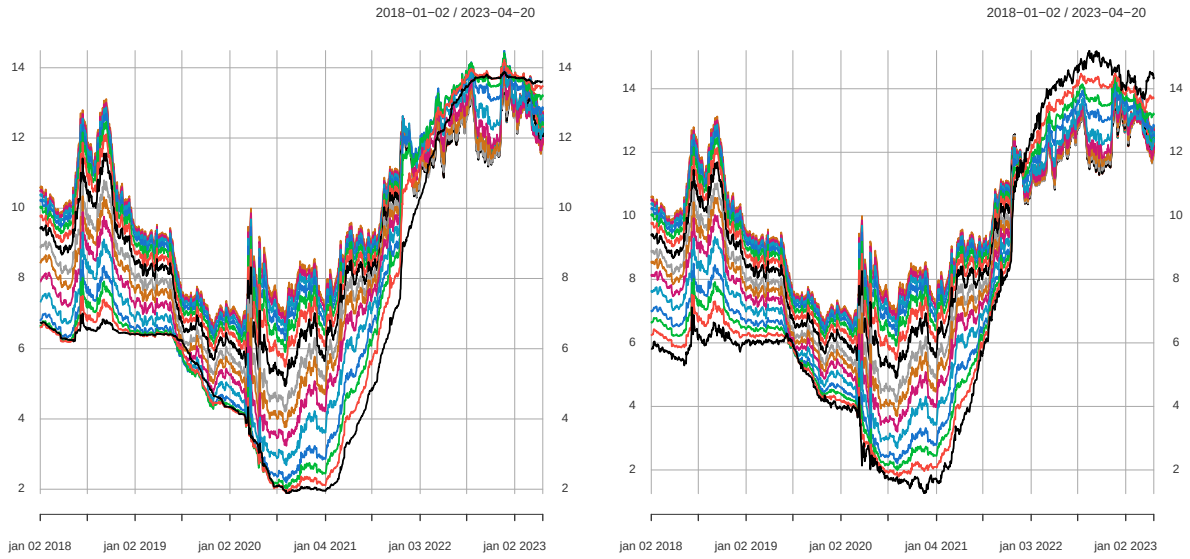
Three series of factor estimates are considered: the data-based factors calculated using yields from the natural cubic splines estimates, the Nelson-Siegel estimates obtained by OLS, and the Nelson-Siegel estimates obtained from the Kalman filter.

The correlations and co-movements of the series support the latent factor interpretation, indicating that the Nelson-Siegel parameters summarise much of the yield curve information. However, the curvature factors estimated by the Kalman filter exhibit a slight deviation from the others, showing a lower correlation.

Table 3.2: Correlation between Model-based and Data-based factors

	Data-based			Model-based (OLS)			Model-based (Kalman)		
	level	slope	curvature	$\beta_{0,\text{OLS}}$	$\beta_{1,\text{OLS}}$	$\beta_{2,\text{OLS}}$	$\beta_{0,\text{KF}}$	$\beta_{1,\text{KF}}$	$\beta_{2,\text{KF}}$
level	1	0.518	0.332	**0.918**	0.572	0.318	**0.981**	0.578	0.078
slope	0.518	1	0.040	0.311	**0.985**	-0.040	0.394	**0.973**	-0.334
curvature	0.332	0.040	1	0.017	0.201	**0.992**	0.237	0.227	**0.577**
$\beta_{0,\text{OLS}}$	0.918	0.311	0.017	1	0.317	0.024	**0.964**	0.314	0.011
$\beta_{1,\text{OLS}}$	0.572	0.985	0.201	0.317	1	0.123	0.439	**0.996**	-0.265
$\beta_{2,\text{OLS}}$	0.318	-0.040	0.992	0.024	0.123	1	0.235	0.149	**0.619**
$\beta_{0,\text{KF}}$	0.981	0.394	0.237	0.964	0.439	0.235	1	0.450	0.023
$\beta_{1,\text{KF}}$	0.578	0.973	0.227	0.314	0.996	0.149	0.450	1	-0.303
$\beta_{2,\text{KF}}$	0.078	-0.334	0.577	0.011	-0.265	0.619	0.023	-0.303	1

Figure 3.2: Kalman filter estimates and actual yields

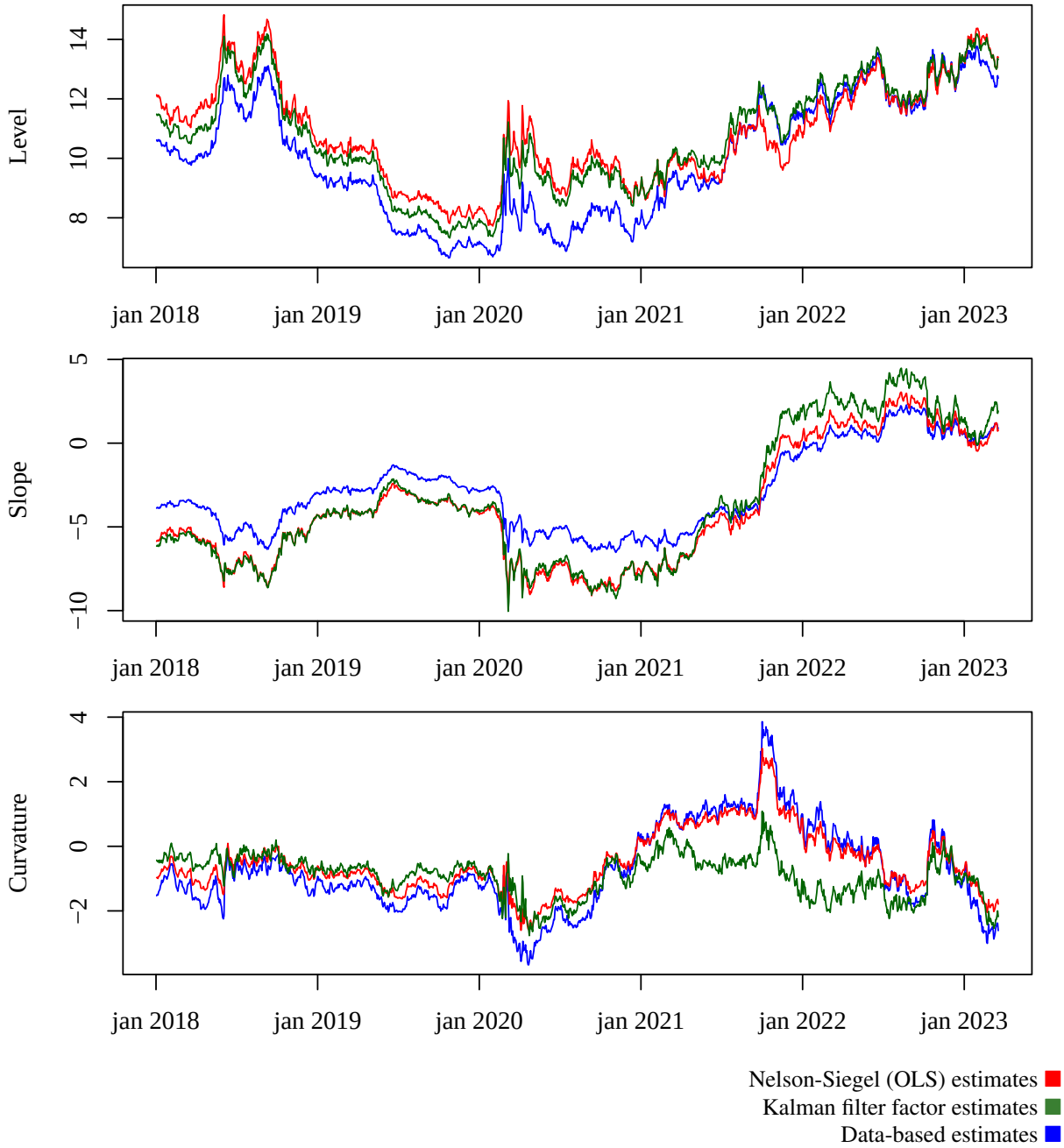


(a) Actual yields evolution

(b) Kalman filter estimates

Source: Elaborated by the author.

Figure 3.3: Model-based and Data-based level, slope, and curvature factors.



Note: Data-based factors correspond to daily measures of: Level $= y_t(120)$, Slope $= -(y_t(120) - y_t(3))$, and Curvature $= \frac{1}{0.3} \times 2y_t(24) - y_t(3) - y_t(120)$; with maturity measured in months.
Source: Elaborated by the author.

3.4.2 Out-of-sample forecasting performance

To evaluate the out-of-sample forecasting performance, we consider a training set with 987 days running from 2 January 2018 to 30 December 2021, representing around 75% of the days in the sample, and a test set with 326 days spanning from 3 January 2022 to 20 April 2023. The period chosen for the test set is particularly interesting for evaluating forecast performance. As shown in figures 2.18 and 3.1, it captures various shapes of the yield curve. It begins as a normal curve, evolves into a humped curve, and eventually transforms into an inverted curve in the subsequent periods.

Four horizons, namely one month, three months, six months, and one year, are employed to evaluate the forecasts. The evaluation of models is conducted through a **walk-forward validation** (or forward chaining) strategy. This approach estimates the models recursively using an expanding window over the test set for each horizon. We define the forecast error h steps ahead at the maturity m as:

$$e_{t+h}(m) \equiv y_{t+h}(m) - \hat{y}_{t+h}(m).$$

For each horizon the Root Mean Square Error (RMSE) at each maturity is presented.

$$RMSE_h(m) = \sqrt{\frac{\sum_{i=m_1}^{M} (e_{t+h}(m))^2}{M}}$$

In Table 3.3 and Table 3.4, RMSE for out-of-sample forecasts are presented across different horizons, namely one-month, three-month, six-month, and twelve-month, spanning from January 2022 to April 2023. The comparison includes forecasts generated by the Random Walk (RW), Exponential Smoothing (HW), Nelson-Siegel modelled as first order auto-regressive (NS (AR)), and Nelson-Siegel modelled as vector auto-regressive (NS (VAR)) models. Notably, the results indicate that both dynamic Nelson-Siegel specifications perform, at best, on par with the random walk forecast across various horizon-maturity scenarios. However, the competitive

models often exhibit similar performance levels, making it challenging to discern a clear superiority among them. Nelson-Siegel specifications perform relatively better at longer maturities.

The RMSE values for the Nelson-Siegel specifications, although higher than those for the random walk in our sample, align with findings from previous research. Specifically, the observed values are consistent with the results reported by Caldeira et al. (2010, p. 45) for three- and six-month forecasts using dynamic Nelson-Siegel estimates with both two-step and one-step estimation procedures.

Table 3.3: RMSE for out-of-sample forecasts (January 2022 - April 2023)

Maturity	One month ahead				Three months ahead			
(months)	RW	HW	NS (AR)	NS (VAR1)	RW	HW	NS (AR)	NS (VAR1)
3	0.360836	0.183001	0.379267	0.297998	0.883602	0.680470	1.310603	0.661419
6	0.311177	0.280051	0.720071	0.417138	0.696030	0.822909	1.541759	0.857052
9	0.364549	0.404355	0.878323	0.612168	0.662001	0.997509	1.612765	1.061739
12	0.461427	0.523285	0.883702	0.685592	0.737891	1.168571	1.554355	1.181116
18	0.628204	0.697699	0.790608	0.733999	0.923757	1.356668	1.355409	1.295441
24	0.704208	0.761764	0.717859	0.762511	1.023426	1.342455	1.198958	1.331268
30	0.725325	0.762786	0.679568	0.783560	1.058070	1.237628	1.093156	1.326141
36	0.727914	0.754846	0.672500	0.796786	1.065510	1.163954	1.047802	1.310544
48	0.722067	0.735294	0.679223	0.780535	1.056660	1.087053	1.076606	1.278453
60	0.711521	0.726293	0.693809	0.757009	1.039495	1.078596	1.136577	1.250975
72	0.699961	0.722628	0.708053	0.743037	1.021190	1.075721	1.178398	1.227135
84	0.687902	0.692448	0.719924	0.734419	1.002296	1.004077	1.206631	1.207037
96	0.677150	0.679353	0.730328	0.729975	0.985348	0.979376	1.226136	1.191272
108	0.669370	0.676366	0.739946	0.728743	0.972728	0.985876	1.240543	1.179641
120	0.664527	0.669290	0.748497	0.729472	0.964690	0.962248	1.251413	1.171194

To evaluate the statistical significance of the difference in the forecasting accuracy of different models, we use the Diebold-Mariano test (Diebold and Mariano, 1995). The null hypothesis is that the two forecasts have the same accuracy, and the alternative hypothesis is that the accuracy is different. The test statistic is based on the comparison of the forecast error (e_{t+h}) produced by the different models (M_1 and M_2).

Let us consider: $d_t \equiv g(e_{t+h,M_1}) - g(e_{t+h,M_2})$, where $g(.)$ is a loss function, and $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$, and $f_d(0)$ the spectral density of d at frequency 0. Under the null hypothesis we

Table 3.4: RMSE for out-of-sample forecasts (January 2022 - April 2023)

Maturity	Six months ahead				Twelve months ahead			
(months)	RW	HW	NS (AR)	NS (VAR1)	RW	HW	NS (AR)	NS (VAR1)
3	1.436782	2.089686	2.427101	1.169927	2.264261	7.127739	4.274294	3.624677
6	1.087949	2.142048	2.558256	1.405076	1.540010	6.978356	4.186049	3.834345
9	0.892931	2.253410	2.529404	1.535254	1.130956	7.002021	3.974402	3.875049
12	0.820218	2.343242	2.377662	1.549841	0.984947	6.819685	3.682832	3.792143
18	0.790369	2.243716	2.015190	1.476108	0.948257	5.778081	3.167886	3.552346
24	0.805346	1.913122	1.743295	1.413426	0.991109	4.448723	2.873178	3.375484
30	0.846843	1.466021	1.568094	1.379319	1.092762	3.088495	2.734377	3.251877
36	0.893568	1.174432	1.487622	1.374473	1.211415	1.967097	2.683500	3.159278
48	0.952551	0.936440	1.508837	1.409003	1.369060	1.242420	2.694599	3.014237
60	0.986320	0.962490	1.579982	1.439781	1.460259	1.139037	2.736700	2.876062
72	1.005815	1.020808	1.625935	1.450027	1.507464	1.293122	2.761343	2.733469
84	1.015293	0.916778	1.653208	1.450493	1.522351	1.087105	2.771152	2.599741
96	1.020174	0.952218	1.668718	1.447241	1.519613	1.273343	2.770277	2.480504
108	1.023477	1.004368	1.678307	1.443441	1.511735	1.395453	2.763599	2.378169
120	1.025995	0.984581	1.684558	1.439923	1.504964	1.241041	2.754163	2.292040

have

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi f_d(0)}{T}}} \sim N(0,1).$$

In this section we use the modified version of the Diebold-Mariano test, following Harvey et al. (1997), which considers the test statistic

$$DM^* = \sqrt{\frac{T+1-2h+h(h+1)}{T}} \times DM \sim t_{T-1},$$

as implemented in the function `dm.test()` from the package `forecast`.

In Table 3.3 and Table 3.4, the DM^* statistic is presented for the different horizons and maturities analysed. Following the interpretation of Caldeira et al. (2010), a negative value indicates the superior accuracy of the first model in the pairwise comparison; in turn, only a statistic equal or greater than 1.96 would be significant at 95%.

The results reveal a nuanced performance across different forecasting horizons. For shorter horizons, the accuracy of the NS(AR) specification is somewhat mixed, and in some instances, it

is outperformed by the random walk (RW). However, the situation is reversed at longer horizons, with the random walk outperforming the NS(AR) specification in all maturities. Interestingly, the random walk and NS(VAR) comparison does not yield significant differences at longer horizons, suggesting a comparable forecasting accuracy between the two models in this context.

Table 3.5: Diebold-Mariano tests (one and three months ahead)

Maturity			One month ahead			
(months)	RWxHW	RWxNS(AR)	RWxNS(VAR)	HWxNS(AR)	HWx_NS(VAR)	NS(AR)xNS(VAR1)
3	1.363532	-0.620856	0.487307	-2.043085	-1.859487	0.792874
6	0.503500	-3.321056	-2.411610	-2.768548	-1.825784	2.789220
9	-0.808256	-2.986749	-2.610254	-2.565919	-1.833769	2.755543
12	-1.159176	-2.514921	-2.422708	-1.955901	-1.392196	2.226347
18	-1.271097	-1.456481	-2.598031	-0.632869	-0.443821	0.743763
24	-1.078024	-0.230946	-2.240613	0.430236	-0.017935	-0.670435
30	-0.926483	1.604127	-1.323625	1.345754	-1.530881	-1.660807
36	-0.956248	2.845019	-1.253002	2.186293	-1.495481	-2.000399
48	-0.697779	2.011436	-1.199975	1.498317	-1.304805	-1.624712
60	-0.852574	0.457769	-1.242215	0.602428	-1.165688	-1.004218
72	-1.256657	-0.163608	-1.438830	0.235673	-1.114168	-0.543807
84	-0.383916	-0.579655	-1.878166	-0.415579	-1.970015	-0.220462
96	-0.235658	-0.906188	-2.506075	-0.754745	-2.607874	0.005269
108	-0.748325	-1.150488	-2.981642	-0.916244	-3.256587	0.164786
120	-0.560092	-1.317765	-3.031053	-1.111725	-3.089192	0.275866
			Three months ahead			
	RWxHW	RWxNS(AR)	RWxNS(VAR)	HWxNS(AR)	HWx_NS(VAR)	NS(AR)xNS(VAR1)
3	0.491688	-2.309746	0.435996	-1.289454	0.067495	1.088589
6	-0.659359	-1.860462	-0.880751	-1.384767	-0.163254	1.273432
9	-2.885082	-1.698736	-3.592556	-1.199837	-0.327287	1.142839
12	-3.011165	-1.478076	-3.781157	-0.755565	-0.056517	0.841884
18	-2.214430	-0.981536	-4.943526	0.002551	0.262365	0.160822
24	-1.541888	-0.550485	-8.372753	0.328029	0.055749	-0.433971
30	-0.989837	-0.157894	-5.620412	0.400549	-0.675585	-0.923887
36	-0.724005	0.109344	-3.374237	0.419098	-2.398539	-1.180889
48	-0.474775	-0.165405	-2.425232	0.058532	-3.371488	-1.004346
60	-0.674301	-0.774401	-2.276211	-0.323517	-3.503083	-0.581752
72	-1.094533	-1.126858	-2.304990	-0.563613	-3.343477	-0.240059
84	-0.031632	-1.332735	-2.395884	-0.995690	-3.626049	-0.001906
96	0.171356	-1.458494	-2.419677	-1.258036	-3.418076	0.156247
108	-0.429979	-1.534405	-2.335896	-1.268294	-3.000458	0.262777
120	0.077274	-1.577922	-2.197719	-1.379453	-2.805802	0.335908

Table 3.6: Diebold-Mariano tests (six and twelve months ahead)

Maturity (months)	RWxHW	RWxNS(AR)	RWxNS(VAR)	HWxNS(AR)	HWx_NS(VAR)	NS(AR)xNS(VAR1)
			Six months ahead			
3	-9.16777523	-1.41830939	0.34470058	-0.47823388	13.1740699	0.96353208
6	-17.5858323	-1.22023776	-5.95909265	-0.58648889	9.70541356	1.05415282
9	-2.89669866	-1.15603669	-1.86134696	-0.4646511	8.16839914	1.08947797
12	-4.01110365	-1.11097847	-1.43287035	-0.06999037	8.02127677	1.08796465
18	-13.770531	-1.0857069	-1.42009104	0.57482543	7.12936813	1.03895229
24	-10.7889661	-1.16625437	-1.97037072	0.51090646	4.83774533	0.96837443
30	-6.43926946	-1.3143259	-3.23402472	-0.37120504	0.93482965	0.92232702
36	-3.61543036	-1.50928781	-4.44204092	-1.02707847	-2.66172948	1.67530886
48	0.39804093	-2.49614098	-10.2408604	-4.99709982	-9.65195086	2.03440247
60	0.79564063	-6.92334231	-10.3589688	-6.19574838	-10.2809476	3.01962859
72	-0.29836431	-19.1052972	-8.93523241	-12.4959222	-2.83171133	3.89502759
84	3.96862461	-19.7807339	-3.41600458	-15.3635863	-10.5149828	4.51744343
96	5.94283549	-20.022551	-2.46012311	-17.7637795	-8.96066568	4.89342926
108	0.66030512	-20.0733263	-2.05541663	-17.6717936	-3.69282774	5.10172322
120	3.9262308	-20.0537387	-1.83511263	-17.5555542	-2.30237404	5.20703454
			Twelve months ahead			
	RWxHW	RWxNS(AR)	RWxNS(VAR)	HWxNS(AR)	HWx_NS(VAR)	NS(AR)xNS(VAR1)
3	-27.28709	-30.0412399	-1.58772243	18.2191472	17.9054074	3.961338
6	-4.85249401	-27.11486	-1.20765896	4.09190464	8.42650135	0.54051077
9	-5.81716019	-23.9990784	-1.07328045	5.27229381	4.86644024	0.11860206
12	-16.9072759	-21.0846129	-0.96267304	17.4207304	2.25335932	-0.05534562
18	-4.48795456	-17.51471	-0.81413235	3.24792565	0.95292642	-0.1953936
24	-1.98124438	-17.0850581	-0.75142177	1.18585134	0.35940382	-0.23036938
30	-1.44685496	-18.3166344	-0.72277152	0.45242457	-0.05467218	-0.2251799
36	-0.47721344	-19.6340178	-0.69594393	-0.93058385	-0.35242166	-0.2027254
48	0.35885891	-14.8391179	-0.64016176	-15.8193292	-0.61717672	-0.13618789
60	1.29642908	-5.1890762	-0.59239511	-24.7145192	-0.62287417	-0.05826171
72	1.6478331	-4.28630113	-0.55023322	-7.53090401	-0.58594189	0.02145237
84	1.65568961	-3.99145882	-0.51496956	-5.10568342	-0.58371264	0.0985989
96	20.053491	-3.86298893	-0.48563972	-4.79631723	-0.55064665	0.17028869
108	6.29638752	-3.7791777	-0.46030249	-4.05205409	-0.50244745	0.23474598
120	16.3238005	-3.70377532	-0.43783354	-4.85783986	-0.51093479	0.2916388

3.5 Conclusion

This chapter comprehensively introduced the dynamic Nelson-Siegel model, exploring various estimation methods. It surveyed the existing literature on the Brazilian yield curve, providing a foundation for the subsequent analysis. The dynamic Nelson-Siegel model was then applied to investigate latent factor interpretations. Parameter estimates were compared across different methodologies, including data-based calculations, daily OLS and Kalman filter estimation of the Nelson-Siegel model. The consistency of these estimates lent support to the latent factor interpretation.

Subsequently, a forecasting exercise was executed, employing the Diebold-Li specification with two-step estimation procedures. The dynamics were modelled as three auto-regression and a vector auto-regression estimation procedures. These results were benchmarked against baseline techniques, such as naive (random walk) and exponential smoothing forecasts.

The dynamic Nelson-Siegel model performed poorly compared with the original results of Diebold and Li (2006) and some previous exercises using Brazilian data. In many instances, it was outperformed by the naive forecast.

Chapter 4

Final Considerations

In conclusion, the two chapters provide valuable insights into yield curve modelling and forecasting aspects. Chapter 2 comprehensively evaluates empirical models for the term structure of interest rates, emphasizing the importance of accurate yield curve estimation in economic analysis. Through a thorough comparison of various models, including spline and parametric approaches, this chapter contributes to understanding the trade-offs involved in model selection. The introduction of Kernel regression and Loess models represents a novel addition to the literature, enhancing the range of available options for yield curve estimation.

Chapter 3 delves into the dynamic Nelson-Siegel model, exploring different estimation methods and assessing latent factor interpretations. The consistent parameter estimates across methodologies support the latent factor interpretation, providing a robust foundation for further analysis. The forecasting exercise, however, reveals the limitations of the dynamic Nelson-Siegel model in comparison to baseline techniques. The model's under-performance, particularly against the naive forecast, highlights the challenges in achieving accurate predictions using this approach.

Taken together, these chapters underscore the complexity of yield curve modelling and the need for careful consideration of model characteristics, estimation methods, and forecasting strategies. While Chapter 1 enriches the tool-kit for yield curve estimation, Chapter 2 sheds

light on the performance of a specific dynamic model in forecasting scenarios. The combined insights contribute to a more nuanced understanding of yield curve dynamics, offering valuable implications for researchers and practitioners in financial modelling and economic analysis.

Appendix A

Math *Addendum*

This appendix lists some useful elementary mathematical results related to the yield curve.

A.1 Continuous compounded interest

Considering a ZCB, the yield-to-maturity (y_t) corresponds to the interest rate that would break even the current bond price (P_t) to its face value (P_T) once the investor holds it over the maturity ($m = T - t$):

$$P_t = \frac{P_T}{(1+y_t)^m} \Leftrightarrow y_t = \left(\frac{P_T}{P_t}\right)^{\frac{1}{m}} - 1.$$

The above formula considers that the interest is calculated once a year, over m years. Interest calculated n times per year results in the formula

$$P_t = \frac{P_T}{\left(\left(1+\frac{y_t}{n}\right)^n\right)^m} \Leftrightarrow \left(1+\frac{y_t}{n}\right)^{n\times m} = \frac{P_T}{P_t}.$$

Continuously compounding the interest rate means that $n \to \infty$, thus we can take the limit on the right-hand-side expression above. Since the yield-to-maturity differs from the discrete

interest calculation we use the notation $\tilde{y}_t$ for the continuous compounded yield-to-maturity.

$$\lim_{n\to\infty}\left(1+\frac{\tilde{y}_t}{n}\right)^{n\times m}=\lim_{n\to\infty}\frac{P_T}{P_t}\Rightarrow\left(\underbrace{\lim_{n\to\infty}\left(1+\frac{\tilde{y}_t}{n}\right)^{n}}_{=e^{\tilde{y}_t}}\right)^{m}=\frac{P_T}{P_t}\Rightarrow e^{m\times\tilde{y}_t}=\frac{P_T}{P_t}$$

$$ln\left(^{m\times\tilde{y}_t}\right)=ln\left(\frac{P_T}{P_t}\right)\Rightarrow m\times\tilde{y}_t=ln\left(\frac{P_T}{P_t}\right)\Rightarrow\boxed{\tilde{y}_t=\frac{ln\left(\frac{P_T}{P_t}\right)}{m}}$$

A.2 Yield-to-maturity, discount function, and implied forward rate

Throughout the main text, we considered the term structure of interest rates expressed in its **yield-to-maturity** (or spot rate) form, $y_t(m)$. This is the logical option given the data, One-day Interbank Deposit Futures ("DI1"), which corresponds to spot rate values. However, other expressions for the term structure of interest rates can be obtained from the relations described below. Let $P_t(m)$ represents the price of a ZCB with maturity m in day t, and let $P_T = 1$ represents bond's face value. If $\tilde{y}_t(m)$ represents the continuously compounded yield-to-maturity, the **discount curve** is given by

$$d_t(m)=e^{-\tilde{y}_t(m)\times m}.$$

The **instantaneous forward rate curve** is given by

$$f_t(m)=-\frac{d_t'(m)}{d_t(m)}.$$

And it relates with yield-to-maturity following

$$\tilde{y}_t(m)=\frac{1}{m}\int_0^m f_t(\nu)d\nu.$$

A.3 Optimal λ for a given maturity in Nelson-Siegel model

In their seminal work, Nelson and Siegel (1987) estimated their model considering a grid of values for the parameter λ in

$$y(m) = \beta_0 + \beta_1 \left(\frac{1 - e^{-m/\lambda}}{m/\lambda} \right) + \beta_2 \underbrace{\left(\frac{1 - e^{-m/\lambda}}{m/\lambda} - e^{-m/\lambda} \right)}_{\text{Curvature loading factor}} + \epsilon_m.$$

For each value assigned to λ, it is possible to estimate the model by ordinary least squares and then choose the appropriate λ value based on a goodness-of-fit criterion. In turn, the dynamic model developed by Diebold and Li (2006, p. 346) used a fixed value for λ, which was justified based on numerical trustworthiness. Considering that λ determines the maturity in which the curvature (or medium-term) loading factor reaches its maximum, they argued that it should occur between two and three years. Thus, they set $1/\lambda = 0.0609$ such that the medium-term loading factor reaches its maximum at precisely 30 months.

We demonstrate how to determine the optimal λ for the medium-term loading factor to reach its maximum at a given maturity. Firstly, one can write the loading factor as a function

$$f(m, \lambda) = \frac{1 - e^{-m/\lambda}}{m/\lambda} - e^{-m/\lambda}.$$

Then consider a fixed value for the maturity, $m = k$, like Diebold and Li (2006) did. Furthermore, consider a variable transformation to simplify the differentiation $v = -k/\lambda$.

$$f(m, \lambda) = \frac{1 - e^{-m/\lambda}}{m/\lambda} - e^{-m/\lambda} \Rightarrow f(\lambda) \quad = \frac{1 - e^{-k/\lambda}}{k/\lambda} - e^{-k/\lambda} \Rightarrow g(v) = \frac{e^v - 1}{v} - e^v$$

Now it is possible to obtain the first-order condition differentiating $g(v)$ with respect to v:

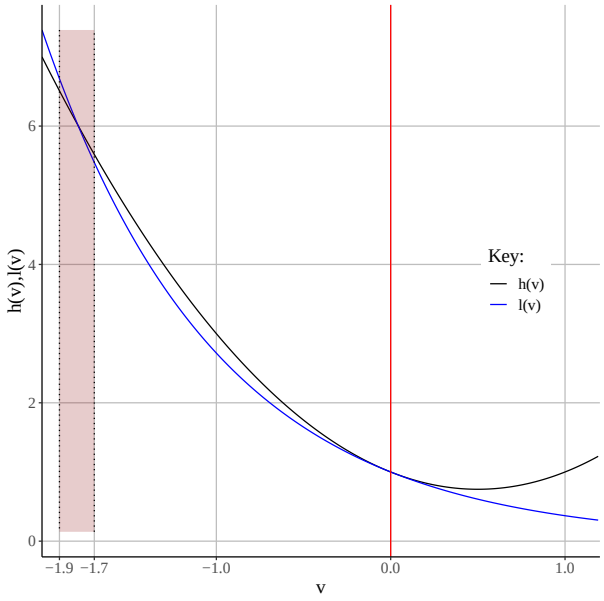
$$\begin{aligned} g'(v) &= \frac{(e^v - 0) \times v - 1 \times (e^v - 1)}{v^2} - e^v \\ &= \frac{ve^v - e^v + 1}{v^2} - e^v = \frac{-v^2e^v + ve^v - e^v + 1}{v^2} \end{aligned}$$

Since the first-order condition for a maximum is $f'(v) = 0$, we can work only with the numerator:

$$-v^2e^v + ve^v - e^v + 1 = 0 \Rightarrow v^2e^v - ve^v + e^v = 1 \Rightarrow \boxed{v^2 - v + 1 = \frac{1}{e^v}}$$

Solving for v will determine the critical points. The trivial solution $v^* = 0$ is not helpful since it implies $\lambda = \infty$. Therefore, one must resource numerical methods to find a different solution. For that, one can write the first-order condition as two functions: $h(v) = v^2 - v + 1$, $l(v) = \frac{1}{e^v}$. Both functions are plotted in Figure A.1.

Figure A.1: Optimal v numerical determination



Source: Elaborated by the author.

```r
v <- seq(-1.85,-1.75, by = 0.00001)
h_v <- v^2 - v + 1
sit1 <- factor("h(v)")
l_v <- 1/exp(v)
sit2 <- factor("l(v)")
numerical2 <- rbind(data.frame(v = v, y = h_v, equation = sit1),
                    data.frame(v = v, y = l_v, equation = sit2))
compare <-
    filter(numerical2, between(v,-1.85,-1.75)) %>%
      pivot_wider(names_from = equation, values_from = y) %>%
      mutate(diff = abs(`h(v)` - `l(v)`))
sprintf("%.100f",compara[which.min(compara$diff),]$v)
```

Listing 1: Optimal v numerical determination

Besides the trivial solution, another critical point lies in the interval [-1.9,-1.7]. The code presented on Listings 1 determines the second critical point numerically: v^* = `-1.79328`. With this value, we can work the transformation $v = -k/\lambda$ backwards to find the optimal λ. Setting k = 30 months, we have

$$v^* = \frac{-k}{\lambda^*} \Rightarrow -1.79328 = \frac{-30}{\lambda^*} \Rightarrow \boxed{\lambda^* = \frac{-30}{-1.79328} \approx 16,72}.$$

Which is close to the value set by Diebold and Li, $1/\lambda = 0.0609 \Leftrightarrow \lambda = 16.42$. Thus, our estimate for λ aligns with Diebold and Li, and differences in the numerical optimization explain the discrepancy.

As mentioned in Chapter 2, λ is not scale-free, and its value depends on how maturity is measured. While Diebold and Li (2006) measured maturity in months, we use working days throughout the main text. Therefore, to reproduce their result, it is necessary to convert the 30 months into working days, approximately $660 (= 30 \times 22)$, to obtain the correspondent λ. Using working days as measurement unit results $\lambda^* = \frac{-k}{v^*} = \frac{-660}{-1.79328} \approx 368$. Alternatively, it is possible to convert it directly: $\lambda^* = 16.72 \times 22 = 367.84$.

Bibliography

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.

Almeida, C., Gomes, R., Leite, A., Simonsen, A., and Vicente, J. (2009). Does curvature enhance forecasting? *International Journal of Theoretical and Applied Finance*, 12(08):1171–1196.

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.

Bowsher, C. and Meeks, R. (2006). High Dimensional Yield Curves: Models and Forecasting. *Economics Papers*. Number: 2006-W12 Publisher: Economics Group, Nuffield College, University of Oxford.

Cajueiro, D., Divino, J. A., and Tabak, B. (2009). Forecasting the Yield Curve for Brazil. Working Papers Series 197, Central Bank of Brazil, Research Department, Brasília, DF.

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., and Portugal, M. S. (2010). Efficient Yield Curve Estimation and Forecasting in Brazil. *EconomiA*, 11(1):27–51.

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.

Campbell, J. Y. and Shiller, R. J. (1991). Yield Spreads and Interest Rate Movements: A Bird's Eye View. *The Review of Economic Studies*, 58(3):495–514.

Carvalho, J. P. D. and Moura, G. V. (2014). Modelo De Fatores Dinâmicos: Estimação E Previsão Da Curva Real De Juros. *Anais do XLI Encontro Nacional de Economia [Proceedings of the 41st Brazilian Economics Meeting]*.

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.

Culbertson, J. M. (1957). The Term Structure of Interest Rates. *The Quarterly Journal of Economics*, 71(4):485–517.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.

Diebold, F. X., Rudebusch, G. D., and Boragan Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1-2):309–338.

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, I. (1896). *Appreciation and Interest*. Macmillan, New York.

Fisher, I. (1930). *The Theory of Interest - As Determined by Impatience to Spend Income and Opportunity to Invest It*. Macmillan, 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.

Goodhart, C. (1989). *Money, Information and Uncertainty*. MIT Press, Cambridge, Mass, second edition.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, New York, NY.

Hayfield, T. and Racine, J. S. (2008). Nonparametric Econometrics: The **np** Package. *Journal of Statistical Software*, 27(5).

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.

Hicks, J. R. (1946). *Value and Capital: An Inquiry into some Fundamental Principles of Economic Theory*. Oxford University Press, London, second edition.

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.

IMF (2018). Brazil - Financial Sector Assessment Program - Technical Note on Systemic Liquidity Management. IMF Country Report 18/345, International Monetary Fund, Washington, D.C.

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.

Koopman, S. J., Mallee, M. I. P., and Van der Wel, M. (2010). Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson–Siegel Model With Time-Varying Parameters. *Journal of Business & Economic Statistics*, 28(3):329–343.

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

Litterman, R. B. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns. *The Journal of Fixed Income*, 1(1):54–61.

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.

Mishkin, F. S. (2000). *Moeda, bancos e mercados financeiros*. Livros Técnicos e Cientifícos Editora, Rio de Janeiro, 5 edition.

Modigliani, F. and Sutch, R. (1966). Innovations in Interest Rate Policy. *The American Economic Review*, 56(1/2):178–197.

Modigliani, F. and Sutch, R. (1969). The Term Structure of Interest Rates: A Reexamination of the Evidence: Reply. *Journal of Money, Credit and Banking*, 1(1):112–120.

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.

Shea, G. S. (1985). Interest Rate Term Structure Estimation with Exponential Splines: A Note. *The Journal of Finance*, 40(1):319–325.

Souza Junior, P. I. F. d. (2021). *Estrutura a termo da taxa de juros no Brasil: Projeções utilizando aprendizado de máquina*. Dissertação (Mestrado em Economia do Setor Público), Universidade de Brasília, Brasília, DF.

Steeley, J. M. (2008). Testing Term Structure Estimation Methods: Evidence from the UK STRIPS Market. *Journal of Money, Credit and Banking*, 40(7):1489–1512.

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

Valentim, J. (2022). *Renda fixa - aplicada ao mercado brasileiro*. Coleção Matemática e Aplicações. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ.

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

Vasicek, O. A. and Fong, H. G. (1982). Term Structure Modeling Using Exponential Splines. *The Journal of Finance*, 37(2):339–348.

Vicente, J. and Tabak, B. M. (2008). Forecasting bond yields in the Brazilian fixed income market. *International Journal of Forecasting*, 24(3):490–497.

Vieira, F., Fernandes, M., and Chague, F. (2017). Forecasting the Brazilian yield curve using forward-looking variables. *International Journal of Forecasting*, 33(1):121–131.

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.