Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

FINANCE

MASTER'S FINAL WORK

DISSERTATION

EU-BONDS YIELD CURVE FORECAST: COMPARING ARIMA AND XGBOOST MODELS

SOFIA ALEXANDRA SANTOS SOARES

JUNE - 2025

Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

FINANCE

MASTER'S FINAL WORK

DISSERTATION

EU-BONDS YIELD CURVE FORECAST: COMPARING ARIMA AND XGBOOST MODELS

SOFIA ALEXANDRA SANTOS SOARES

SUPERVISION:

RAQUEL M. GASPAR
ANA PARDAL

JUNE - 2025

Acknowledgements

I wish to thank my supervisors, Professor Raquel Gaspar and Ana Pardal, for their valuable guidance and suggestions that allowed me to successfully write this dissertation.

I am thankful to the G10 Rates Sales team at BNP Paribas for giving me a unique opportunity to integrate theoretical knowledge into a professional setting, providing a good work environment and all the necessary data for my thesis.

I also want to thank my friends and my colleagues at BNP Paribas for their help while I was pursuing this project.

Last but not least, I want to express my gratitude to my family for the unwavering support provided during all my academic path.

ABSTRACT

The goal of this dissertation is to model and forecast the yield curve of the bonds issued by the European Union, the EU-Bonds, using 2024 daily yield data for this product.

Two models are used and compared for modelling the yield curve: the Nelson-Siegel model and the Svensson model. These models are calibrated daily during the year of 2024.

A classic time series econometric model, the ARIMA, and a Machine-Learning model, the XGBoost, are used to forecast the yield curve. The forecast is performed daily for January 2025. The results presented focus on the forecasted yield curves for 1 day, 1 week (7^{th} of January) and 1 month (31^{st} of January) forecast horizons.

The MSE, RMSE, MAE, and MAPE are the error metrics used to evaluate the forecasting performance of these models.

For the 1 day forecast horizon, the models that present the best results are the Svensson model with the XGBoost forecasted parameters, according to the MSE and RMSE metrics, and the Svensson model with the ARIMA predicted parameters, according to the MAE and MAPE metrics.

For the 1 week and 1 month forecast horizons, the models that produce the best results are the Svensson model with the forecasted parameters obtained with ARIMA, according to all error metrics used.

It is concluded that the XGBoost performs better in smaller windows of time and ARIMA has a better forecast accuracy for larger forecast horizons.

KEYWORDS: Bonds; Forecasting; Nelson-Siegel; Svensson; ARIMA; XGBoost.

JEL CODES: C22; C45; C51; C53; E43; G12; G17.

RESUMO

O objetivo desta dissertação consiste em modelar e prever a curva de rendimento das obrigações emitidas pela União Europeia, as "EU-Bonds", usando as taxas de juro diárias de 2024 deste produto.

Dois modelos são usados e comparados para modelar a curva de rendimento: o modelo de Nelson-Siegel e o modelo de Svensson. Estes modelos são calibrados diariamente durante o ano de 2024.

Um modelo econométrico clássico de séries temporais, o ARIMA, e um modelo de "Machine-Learning", o XGBoost, são utilizados para prever a curva de rendimento. A previsão é realizada diariamente para janeiro de 2025. Os resultados apresentados focam-se nas curvas de rendimento previstas para os horizontes temporais de 1 dia, 1 semana (7 de janeiro) e 1 mês (31 de janeiro).

O erro ao quadrado médio, a raíz quadrada do erro ao quadrado médio, o erro absoluto médio e o erro percentual absoluto médio são as métricas de erro usadas para avaliar a performance preditiva destes modelos.

Para a previsão com um horizonte temporal de 1 dia, os modelos que apresentam os melhores resultados são o modelo de Svensson com os parâmetros previstos com o XGBoost, de acordo com as métricas do erro ao quadrado médio e da raíz quadrada do erro ao quadrado médio, e o modelo de Svensson com os parâmetros previstos com o ARIMA, de acordo com o erro absoluto médio e o erro absoluto médio percentual.

Para as previsões com horizontes temporais de 1 semana e 1 mês, os modelos que produzem os melhores resultados são o modelo de Svensson com os parâmetros previstos obtidos com o ARIMA, de acordo com todas as métricas de erro usadas.

É concluído que o XGBoost tem uma melhor performance em janelas de tempo mais reduzidas e o ARIMA tem uma melhor precisão de previsão em horizontes temporais maiores.

PALAVRAS-CHAVE: Obrigações; Previsão; Nelson-Siegel; Svensson; ARIMA; XG-Boost.

CÓDIGOS JEL: C22; C45; C51; C53; E43; G12; G17.

GLOSSARY

ACF Autocorrelation Function. 21

ADF Augmented Dickey-Fuller. 12, 13, 21

AI Artificial Intelligence. 6

AIC Akaike Information Criterion. 12, 13

ARIMA Autoregressive Integrated Moving Average. 1, 6, 7, 11, 12, 16, 22, 23, 29–32

ARMA Autoregressive Moving Average. 7

ESS Explained Sum of Squares. 11

EU European Union. 1–4, 8, 32, 33

GDP Gross Domestic Product. 33

GNI Gross National Income. 2

MAE Mean Absolute Error. 16, 25, 28–30, 32

MAPE Mean Absolute Percentage Error. 16, 25, 28–30, 32

MFA Macro-Financial Assistance. 2

ML Machine-Learning. 1, 6, 7, 11, 13, 33

MSE Mean Squared Error. 11, 15, 16, 18, 19, 25, 28–30, 32

NGEU NextGenerationEU. 2, 3

RMSE Root Mean Squared Error. 16, 25, 28–30, 32

RRF Recovery and Resilience Facility. 3

RRPs Recovery and Resilience Plans. 3

RSS Residual Sum of Squares. 11, 13

SURE Support to mitigate Unemployment Risks in Emergency. 2, 3

TSS Total Sum of Squares. 11

XGBoost eXtreme Gradient Boosting. 1, 7, 11, 13–16, 22–24, 29–33

Table of Contents

List of Figures

List of Tables

1 Introduction

In this dissertation, the European Union (EU)-Bonds yield curve is forecasted for January 2025. This is the chosen product because it is recent in the market and is gaining a lot of importance amongst the market participants. Therefore, it is useful to have access to yield curve forecasts of this product, as it can create lucrative trade ideas. This is especially important since Euro-EU Bond Futures will soon start to be traded, and their underlying are EU-Bonds.

To date, there are not many studies that explore the characteristics of this product and, to the best of my knowledge, this is the first study that intends to predict the EU-Bonds yield curve.

To forecast the EU-Bonds yield curve for January 2025, two models are used: Autoregressive Integrated Moving Average (ARIMA), a classic time series econometric model, and eXtreme Gradient Boosting (XGBoost), a Machine-Learning (ML) model, based on a dataset containing daily EU-Bonds yield values for a diverse set of maturities during the entire year of 2024. The focus of this forecasting exercise is to present the forecasting results for three points in time within the 1 month time frame. This way, we can conclude which model has a better forecasting performance in each scenario.

This document is organised into six chapters. Chapter 2 - EU-Bills and EU-Bonds gives a historical overview of the issuance of EU-Bills and EU-Bonds. In Chapter 3 - Literature Review, a brief description of different models for fitting and predicting the yield curve is presented. Chapter 4 - Data and Methodology presents and explains the models used for calibrating the yield curve, namely the Nelson-Siegel and the Nelson-Siegel-Svensson model (hereinafter referred to as the Svensson model), and the models used for forecasting the yield curve, specifically the ARIMA and the XGBoost. In this chapter, it is also described the dataset used in this study and its descriptive statistics. In Chapter 5 - Results, the fitting and forecasting results of the EU-Bonds yield curves are presented and discussed. Fitting and forecasting models are compared to assess which models are the best for 3 different forecasting horizons: 1 day, 1 week and 1 month. In Chapter 6 - Conclusion, the main findings and limitations of this work are described. Suggestions for future research are also provided.

2 EU-Bills and EU-Bonds

In this chapter, we provide more information regarding the framework behind the issuance of EU-Bills and EU-Bonds. This information can be found on the European Commission and European Council's websites.

The EU has been issuing debt for the past 40 years. Nowadays, it has 7 policy programmes that are funded by the EU debt issuance, exclusively denominated in euro. These programs are: Balance of Payments, Euratom, European Stabilisation Mechanism, Macro-Financial Assistance (MFA), NextGenerationEU (NGEU), Support to mitigate Unemployment Risks in Emergency (SURE) and Ukraine Facility.

The EU-Bonds are guaranteed by the EU Budget, which is financed through three different sources of funds: own resources, surplus of the EU revenue, and other sources, which include fines, refunds and taxes on salaries. The own resources represent more than 90% of the budget and they are mainly funded by contributions of all Member States in proportion to their Gross National Income (GNI). Own resources also include customs duties and contributions based on value-added tax and non-recycled plastic packaging waste. We can infer from this information that the EU Budget is mostly the result of each Member State's contribution in proportion to their GNI. Since the EU Budget is the guarantee of the EU-Bonds, the EU default risk is concentrated on the risk behind the EU Budget. This means that the risk of an EU-Bond should reflect, in theory, a weighted average of the default risk of the EU countries.

The EU funding instruments are EU-Bills and EU-Bonds. The EU-Bills are financial instruments issued since September 2021 for maturities of less than one year, through auctions. The EU-Bonds are issued for maturities of 3, 5, 7, 10, 15, 20, 25 and 30 years, through syndications and auctions. The European Commission can also resort to private placements in specific cases. It has been mostly used for the MFA programme. EU-Bonds can be issued via "taps", which means the European Commission issues more volume of a previously issued bond, or by creating an entirely new bond. More recently, the EU issued two different EU-Bonds: Green Bonds and Social Bonds. The former were issued under the NGEU programme, and the latter under the SURE programme. The SURE and NGEU programmes were created in response to the COVID-19 crisis.

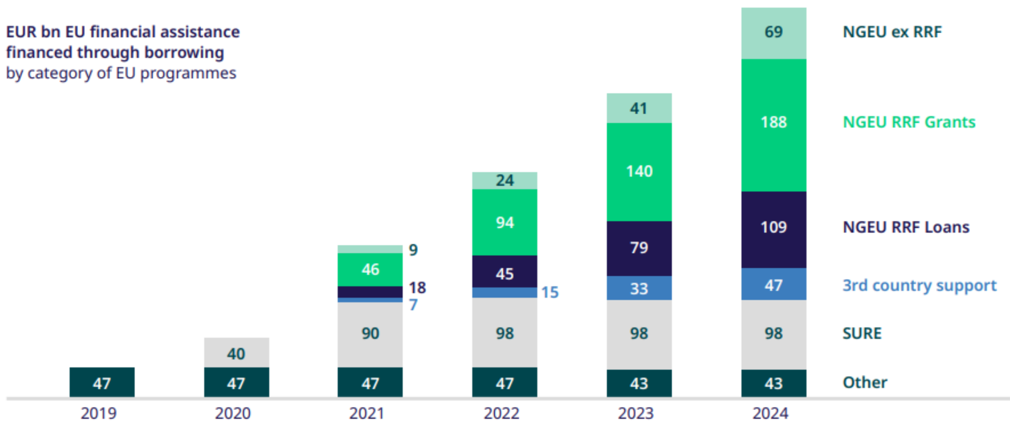
One the one hand, SURE Social Bonds were aimed to aid countries regarding employment issues derived from this pandemic. The issuance of these bonds occurred between October 2020 and December 2022 and raised €98.4 billion. The disbursements made to the EU countries were in the form of loans. Italy, Spain and Poland were the three countries that benefited most from SURE, receiving €27.44 billion, €21.32 billion and

€11.24 billion, respectively.

On the other hand, from the total €806.9 billion NGEU funds, almost 90% were allocated to the Recovery and Resilience Facility (RRF) both in the form of grants (€338.0 billion) and loans (€385.8 billion) and the rest was allocated to other projects. The RRF was created in February 2021, also in response to the pandemic crisis. The proceeds from this instrument are distributed to the EU Member States to help them achieve the proposed improvements on their submitted Recovery and Resilience Plans (RRPs), with a focus on the green and digital transitions. The securities issued under the NGEU programme were first issued in June 2021. Green Bonds were created to finance sustainable and green projects presented in the Member States' RRPs, and the European Commission's goal is that they represent 30% of the total NGEU funds. The first Green Bonds were issued in October 2021. The three countries that received the highest RRF funding values were Italy, Spain and Poland. They received €194 billion, €163 billion and €60 billion, respectively. Looking specifically at NGEU Green Bonds, the countries entitled to the highest amounts of proceeds are Italy (€75 billion), Spain (€68 billion) and Poland (€26 billion).

In Figure 1, we can see how the issuance of bonds under the SURE and NGEU programmes has become increasingly important in the EU issuance.

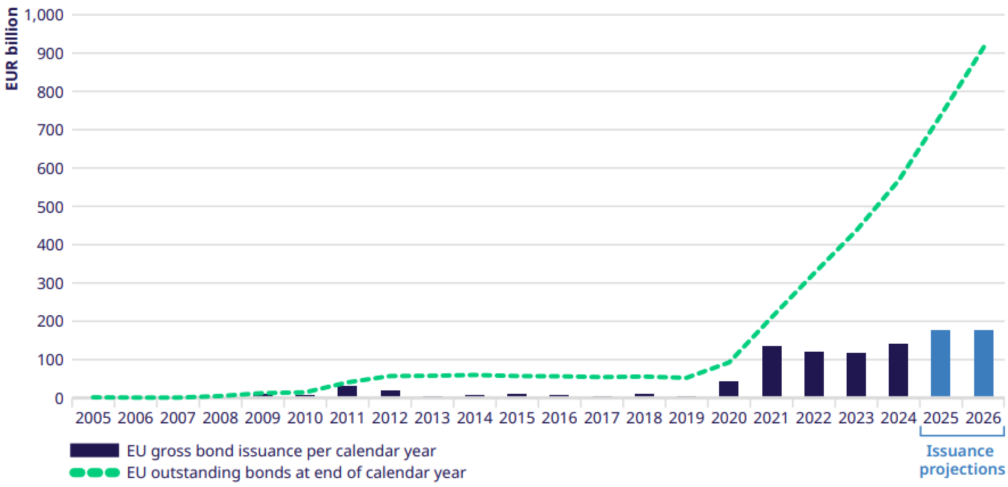
Figure 1: EU Funded Programmes



Source: Eurex (2025)

According to the European Parliament, Policy Department for Budgetary Affairs (2023), "Of the approximately 400 EUR billion in outstanding EU debt as of May 2023, 85% arises from borrowing since 2020.". The SURE and NGEU programmes are the principal cause of this transformation in the EU issuance trajectory, represented in Figure 2. The introduction of these two programmes considerably increased the total amount of bonds outstanding issued by the European Comission, as “Over 93% of the bonds issued by the Commission between October 2020 and December 2022 went towards financing these two instruments.”.

Figure 2: Evolution of EU Issuance



Source: Eurex (2025)

Since January 2023, the European Commission has decided that all bonds would be named as EU-Bonds, not having a reference to the specific programme they are funding. This is called the unified funding approach. In this framework, the European Commission works with a group of banks that form the Primary Dealer Network. This network is responsible for executing syndications and auctions. In this way, the EU-Bonds can be more easily accessed by a large number of investors.

According to Eurex (2025), on the 10th of September of 2025, the Euro-EU Bond Futures contracts will be launched. They will provide a more complete market for the EU-Bonds as they will add the derivatives component to the cash and repo components already existing. This derivative will also provide more liquidity and cost efficiency to the market, creating the possibility to generate accurate hedging positions for investors who possess EU-bonds, as well as spread and basis trading opportunities. These futures contracts include the physical delivery of the underlying, EU-Bonds with maturities between 8 to 12 years, and a 6% coupon, like other Eurex’s fixed income futures with a similar time to maturity.

Various models were created throughout the years to describe the term structure of interest rates. Next, the main characteristics of the different models for calibrating the yield curve are presented.

In McCulloch (1975), the yield curve is parametric and derived from a discount function that is estimated using regression cubic splines, which are piecewise cubic polynomials joined at knot-points. This discount function is constructed with the observed prices of securities with different maturities. Using a cubic spline to fit this function provides more flexibility.

The Vasicek model presented in Vasicek (1977) is one of the first models created to represent the term structure of interest rates. It is an equilibrium model in which the spot rate is given by a stochastic differential equation. The Cox-Ingersoll-Ross model described in Cox et al. (1985) uses an equation similar to the Vasicek model but excludes the possibility of negative interest rates. Both these models are stochastic in nature.

Later, a new parametric model was created by Nelson & Siegel (1987). This is a model that defines the term structure of the yield curve using a parsimonious model with four parameters, β_0, β_1, β_2, and τ. β_0, β_1, and β_2 have an economic meaning and represent the level, slope and curvature of the yield curve, respectively. The Nelson-Siegel model is described in more detail in Section 4.2.1.

An extension of this model is introduced by Svensson (1994). This development introduces two new parameters, β_3 and τ_2. This adds a second curvature to the yield curve and, consequently, fits longer maturities better than the previous model. This makes it more flexible to fitting different term structures. According to the ECB (2023), this model is used for estimating their daily yield curves. The Svensson model is explained more thoroughly in Section 4.2.2.

In Fisher et al. (1995), a new method was introduced to compute the discount function also using splines. In this case, the authors created the smoothing splines method, which does not imply a specific number of parameters as the cubic spline method mentioned above. This model uses smoothing splines that penalise roughness. This means that it penalises the presence of parameters that are not significant in explaining the discount function and, thus, not provide a good fit to the data. If the penalty increases, the number of parameters decreases.

Until that point, the existing models had focused mainly on modelling the yield curve and not on forecasting it. In Diebold & Li (2006), the forecasting of yield curves is addressed in the dynamic Nelson-Siegel model. It considers that the parameters of the Nelson-Siegel model that represent the level, slope and curvature of the yield curve are time-varying and chooses to fix the fourth parameter, τ. They use autoregressive models to forecast the out-of-sample values of the time-varying parameters for short and long-term horizons. Then, the forecasted parameters are used to forecast the yield curve.

In Christensen et al. (2007), arbitrage-free versions of the dynamic Nelson-Siegel model are presented. They concluded that imposing a restriction on the existence of arbitrage yields better results in out-of-sample forecasting, especially for longer maturities and longer forecasting windows.

Christensen et al. (2009) presents a dynamic version of the Svensson model. This model does not imply no arbitrage. To solve this issue, a dynamic generalized Nelson-Siegel model is derived as well as its arbitrage-free version. The dynamic generalized Nelson-Siegel model includes one level, two slopes, and two curvature parameters. It is necessary to add a second slope factor because, for an arbitrage-free model, it is not possible to assign values to two curvature factors with just one slope parameter. They concluded that the arbitrage-free version of the generalized Nelson-Siegel model is tractable and fits well the yield curve.

According to BIS (2005), most central banks have been using the Nelson-Siegel or Svensson models to calculate their yield curves. Countries like Canada, Japan, Sweden (partially), the United Kingdom, and the United States use a version of the smoothing splines method.

More recently, ML models have been used for predicting yield curves. In Castellani & dos Santos (2006), the authors compare the forecasting results of different artificial intelligence and classic models: fuzzy logic, self-organising map, multi-layer perceptron, ARIMA, and error correction model. The latter presents the best results, followed by the multi-layer perceptron and the ARIMA models. The fuzzy logic and the self-organising map gave the poorest results. This paper concludes that Artificial Intelligence (AI) models can capture the main trend of the yield, but probably could not give best results due to the lack of data available to train those models. The classical models, ARIMA and the error correction model, gave results similar to a one-step lagged system. This means using the previous data point to predict the next one. The authors state that combining statistical or machine learning models with expertise in the field can give the best results when forecasting the yield curve.

Gaussian Processes are presented in Sambasivan & Das (2017) as a statistical ML approach to forecast the yield curve. The best results are obtained in the medium and long time to maturity parts of the curve.

Common functional principal component analysis is used to model the yield curve of several economies in Zhang et al. (2017). An AR(1) process, a first-order autoregressive model, is performed to forecast the single common factors. The predicted yield curve results directly from these forecasted factors.

A deep learning Nelson-Siegel model is presented in Lee (2023). This model is an extension of the previous dynamic Nelson-Siegel model developed by Diebold & Li (2006). The author concluded that the deep learning Nelson-Siegel model outperforms the dynamic Nelson-Siegel model in the 1, 3, 6 and 12 months out-of-sample forecasting windows.

In Jeaab et al. (2024), three ML models have been used to predict Morocco's yield curve: long- and short-term memory, gated recurrent units and XGBoost. The results show that the XGBoost model is the best amongst the three to forecast the yield curve.

Autorregressive and ML models are used in Castello & Resta (2024) to predict yield values with a 1 day forecasting horizon. In this paper, two different models for fitting the yield curve are used: the 3-factor dynamic Nelson-Siegel model, and the 5-factor dynamic De Rezende-Ferreira model, presented in Rezende & Ferreira (2008). Its parameters are forecasted using 3 different methods: a univariate autoregressive process, AR(1), also utilised in Diebold & Li (2006), the trigonometric seasonal Box–Cox transformation with Autoregressive Moving Average (ARMA) residuals trend and seasonal components, and ARIMA combined with a non-linear autoregressive neural network. For in-sample fitting, the best results were obtained with time-varying decay parameters, which is done as well in this work, as described in Section 4.2.3. The 5-factor model exhibited better fitting to the data, since it is more flexible than the 3-factor model because it contains more parameters. The general out-of-sample forecasting results were good, with an average predictive precision of more than 95%. The 3-factor dynamic Nelson-Siegel model, combined with the AR(1) process, registered the best results. This comes to show that having more complex models does not always translate into better forecasting results.

In BIS (2021), an analysis of big data and ML in central banking is conducted, and the results of a survey on the Irving Fischer Committee members are presented. It is concluded that central banks can benefit from using big data and ML, helping them to accomplish their mandates. However, some difficulties arise when it comes to obtain, manage and analyse big data. Other concerns are regarding cybersecurity and legal aspects. Budget constraints and the training and hiring of professionals are also issues that prevent central banks from taking full advantage of the possibilities of big data and ML. The results of the survey show that central banks are interested in collaborating together to unlock the full potential of these tools.

4 DATA AND METHODOLOGY

4.1 Data Selection and Descriptive Statistics

The dataset is composed of EU-Bonds daily yields (mid yield to maturity[1]) retrieved from Bloomberg for the period between the 1st of January 2024 and the 31st of January 2025, including fifteen different maturities: 0.25, 0.5, 1, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 and 30 years. This is the time frame chosen due to constraints in the availability of data. The data is divided into training and testing sets. The training set, used for calibrating the yield curve and training the forecast models, contains yields from the 1st of January 2024 until the 31st of December 2024. The testing set, used for comparing the forecasted values with the actual values, includes data from the 1st of January 2025 until the 31st of January 2025.

In Table I, the descriptive statistics of the dataset are presented. The mean of EU-Bill yields with 3 months maturity (0.25 years), 3.4446%, is higher than the mean of EU-Bond yields with 30 years maturity, 3.3464%. The mean of EU-Bill yields with 6 months maturity (0.50 years), 3.3347%, is also higher than the mean yields of other longer-term bonds. This indicates the presence of an inverted yield curve. As expected, the volatility of short-maturity rates, given by the value of the standard deviation, is much larger (almost four times higher) than long-maturity rates.

Table I: Descriptive Statistics of the Dataset

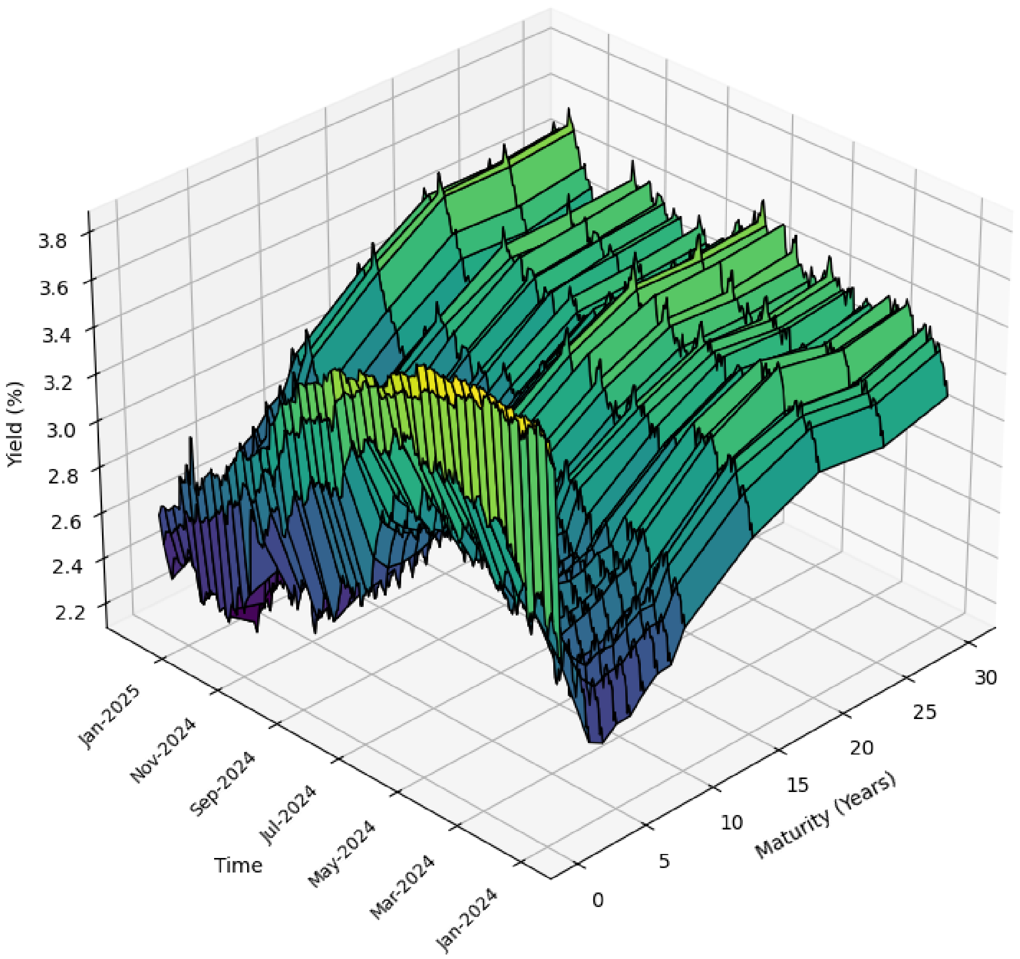
Maturity (Years)	Mean	Std. dev.	Minimum	Q1	Q2 (Median)	Q3	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(7)$	$\hat{\rho}(30)$
0.25	3.4446	0.3975	2.5920	3.0668	3.6180	3.8090	3.8490	0.9978	0.9923	0.9729
0.50	3.3347	0.4308	2.4560	2.9142	3.5285	3.7100	3.8170	0.9989	0.9941	0.9768
1	2.9770	0.3908	2.2330	2.7370	3.0605	3.3285	3.4760	0.9963	0.9695	0.8517
3	2.6962	0.2985	2.1270	2.4420	2.6610	2.9878	3.1980	0.9900	0.9332	0.7362
4	2.6751	0.2248	2.1920	2.4982	2.6325	2.8785	3.1080	0.9819	0.8822	0.6139
5	2.6925	0.1892	2.2620	2.5512	2.6745	2.8598	3.0650	0.9744	0.8376	0.5288
6	2.7121	0.1605	2.3070	2.6012	2.7160	2.8448	3.0360	0.9655	0.7722	0.3887
7	2.7338	0.1578	2.3490	2.6235	2.7465	2.8625	3.0640	0.9652	0.7673	0.3693
8	2.7855	0.1384	2.4450	2.6850	2.8065	2.8928	3.0710	0.9563	0.7209	0.2614
9	2.8107	0.1328	2.4770	2.7170	2.8295	2.9080	3.1150	0.9534	0.6964	0.1778
10	2.9325	0.1267	2.6050	2.8385	2.9405	3.0290	3.2270	0.9495	0.6727	0.0914
15	3.1603	0.1128	2.8410	3.0855	3.1690	3.2425	3.4370	0.9402	0.6221	−0.0679
20	3.3183	0.1028	3.0160	3.2550	3.3265	3.3848	3.5730	0.9317	0.5600	−0.2701
25	3.3068	0.1020	3.0250	3.2438	3.3165	3.3658	3.5640	0.9319	0.5893	−0.3162
30	3.3464	0.0957	3.0550	3.2892	3.3530	3.4030	3.5920	0.9248	0.5585	−0.3839

In Figure 3, we can see a plot of the daily yield curves throughout the selected time

[1] Mid yield to maturity refers to the mid yield between the bid and ask yields.

window before the fitting process, explained in Section 4.2. We can see that the short-term yields are higher than medium and long-term yields for most of the year of 2024, confirming the existence of an inverted yield curve, as shown by Table I.

Figure 3: EU-Bonds Daily Yield Curves



EU-Bonds daily yield curves at maturities of 0.25, 0.5, 1, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 and 30 years, from January 2024 until the end of January 2025.

4.2 Calibrating the Yield Curve

4.2.1 Nelson-Siegel Model

Nelson & Siegel (1987) model is a parametric parsimonious model that describes the yield curve's shape. The spot rate is given by the following formula,

$$s(m) = \beta_0 + \beta_1 \frac{1 - e^{-\frac{m}{\tau}}}{\frac{m}{\tau}} + \beta_2 \left(\frac{1 - e^{-\frac{m}{\tau}}}{\frac{m}{\tau}} - e^{-\frac{m}{\tau}} \right) \tag{1}$$

where $s(m)$ is the spot rate for maturity m, and β_0, β_1, β_2, and τ are the parameters of the model, which have different interpretations.

β_0 is the level factor and is related to the long-term yields. It limits the level of the spot rate as m goes to infinity and is strictly positive. β_1 defines the slope of the curve, influencing the short-term rate. β_2 determines the curvature of the curve, being related to the medium-term rate. τ is the exponential decay rate, in years to maturity, of the slope and curvature factors. This parameter controls where the curvature is located and is strictly positive.

4.2.2 Svensson Model

Svensson (1994) model is an extension of the previous one. It introduces two new parameters that give more flexibility to the model, β_3 and τ_2,

$$s(m) = \beta_0 + \beta_1 \frac{1 - e^{-\frac{m}{\tau_1}}}{\frac{m}{\tau_1}} + \beta_2 \left(\frac{1 - e^{-\frac{m}{\tau_1}}}{\frac{m}{\tau_1}} - e^{-\frac{m}{\tau_1}} \right) + \beta_3 \left(\frac{1 - e^{-\frac{m}{\tau_2}}}{\frac{m}{\tau_2}} - e^{-\frac{m}{\tau_2}} \right) \tag{2}$$

β_3 adds a second curvature to the yield curve, and τ_2 ($\tau_2 > 0$) determines the location of this second curvature.

4.2.3 Fitted Parameters

The previous models are used to fit the data of the training set, obtaining daily yield curves. With this, the parameters for the Nelson-Siegel and the Svensson models, $\theta_{NS_t} = \{\beta_{NS0_t}, \beta_{NS1_t}, \beta_{NS2_t}, \tau_{NS_t}\}$, and $\theta_{S_t} = \{\beta_{S0_t}, \beta_{S1_t}, \beta_{S2_t}, \beta_{S3_t}, \tau_{S1_t}, \tau_{S2_t}\}$, respectively, are estimated for each day t. Consequently, we obtain the following time series: $\{\hat{\beta}_{NS0_t}, \hat{\beta}_{NS1_t}, \hat{\beta}_{NS2_t}, \hat{\tau}_{NS_t}\}$ and $\{\hat{\beta}_{S0_t}, \hat{\beta}_{S1_t}, \hat{\beta}_{S2_t}, \hat{\beta}_{S3_t}, \hat{\tau}_{S1_t}, \hat{\tau}_{S2_t}\}$. The calibration of the yield curve and the estimation of parameters are performed using the non-linear least squares method in Python.

First, it is necessary to calibrate the Nelson-Siegel and Svensson models so we can obtain the time series of the parameters that are forecasted using the models presented in Section 4.3. Then, the forecasted parameters are used as inputs in Equations (1) and (2) to forecast the January 2025 yield curves.

4.2.4 Evaluating the Goodness of Fit

The measures used to evaluate if the Nelson-Siegel and Svensson models provide a good fit to the 2024 yield data are the Mean Squared Error (MSE) and the Adjusted R^2. We perform these calculations to check if the models provide good results, so they can be used to calculate the necessary time series of parameters.

The MSE represents the mean of the squared difference between the actual and the forecasted values, at each point in time i. A lower MSE value indicates a better fit to the data.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\text{Actual}_i - \text{Forecasted}_i)^2 \tag{3}$$

The Coefficient of Determination, R^2, is the proportion of the Explained Sum of Squares (ESS) in the Total Sum of Squares (TSS) or, equivalently, one minus the proportion of the Residual Sum of Squares (RSS) in the TSS. The values can vary between 0 and 1. The higher the value, the better the model fits the data. This measure is suited for evaluating linear regression models (Esaki 2021), which is why it is not used in Section 4.3.4 because one of the evaluated forecasting models, XGBoost, is non-linear.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(\text{Actual}_i - \text{Forecasted}_i)^2}{\sum_{i=1}^{n}(\text{Actual}_i - \bar{\text{Actual}})^2} \tag{4}$$

The Adjusted R^2, denoted as R^2_{adj}, is similar to the previous measure but incorporates the number of observations, n, and a penalization when more predictors that do not explain the dependent variable are added. Consequently, it is a more accurate measure when comparing Nelson-Siegel and Svensson models which have a different number of predictors or parameters, p.

$$R^2_{\text{adj}} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \tag{5}$$

4.3 Forecasting the Yield Curve

Two methods are used for forecasting the time series of each parameter. The first one is a classic time series econometric model, the ARIMA. The second one used is XGBoost, a ML model.

4.3.1 ARIMA Model

ARIMA is a model introduced by Box & Jenkins (1970) that can be used for analysing and forecasting non-stationary and seasonal time series. A non-stationary process is characterized by having varying descriptive statistics, such as mean and variance. A seasonal time series shows specific trends in certain periods of the year. This model can transform non-stationary into stationary time series and seasonal into non-seasonal time series. These are necessary conditions to perform analyses and forecasts well. This study focuses on non-stationary and non-seasonal time series.

The general ARIMA(p, d, q) model is described by the following formula in Box et al. (2016):

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \tag{6}$$

where

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\end{aligned}$$

- $\phi(B)$ is a stationary autorregressive operator of order p;
- $\theta(B)$ is an invertible moving average operator of order q;
- $\varphi(B) = \phi(B)\nabla^d$ is a non-stationary generalized autorregressive operator, where $\nabla^d = (1 - B)^d$ is the differencing operator of order d;
- z_t represents the observations of a given time series in moment t, which in this case is day t;
- θ_0 is a constant term;
- a_t represents a white noise process, which is a stationary process of independent and identically distributed random variables that are assumed to have mean zero and variance σ_a^2.

With this model, it is possible to forecast non-stationary data because of the presence of the differencing operator. This operator differentiates the time series, transforming it into a stationary process. The order d is determined by the number of times necessary to differentiate the data to achieve a stationary time series.

The best orders (p, d, q) for ARIMA are found through the usage of the package *pmdarima* in Python. To choose d, it performs the Augmented Dickey-Fuller (ADF) test. Then, it uses the Akaike Information Criterion (AIC) to select the best ARIMA model.

AIC is a criterion that helps choose the best model and it was introduced by Akaike (1974):

$$\text{AIC} = -2 \log L + 2m \tag{7}$$

where L represents the maximum likelihood function and m is the number of parameters of the model. The lower the value of AIC, the better the model fits to the data.

The ADF is a statistical test that is used to check if a time series is stationary or not. The null hypothesis states that there is a unit root (the process is non-stationary), and the alternative hypothesis states the opposite (the process is stationary).

4.3.2 XGBoost Model

XGBoost is a ML model developed by Chen & Guestrin (2016) and consists of a scalable tree boosting system. It is based on a simpler ML model, called a decision tree. XGBoost uses an ensemble method called gradient boosting that combines various decision trees to get more accurate results. This model is implemented in Python using the package *xgboost*.

There are two types of decision trees: regression and classification trees. XGBoost uses regression trees.

According to James et al. (2023), these trees are constructed through recursive binary splitting of all the observations. First, a predictor X_j and a cutpoint s need to be selected in order to divide the predictor space into two regions, R_1 and R_2, defined as

$$\begin{aligned} R_1(j,s) &= \{X \mid X_j < s\} \\ R_2(j,s) &= \{X \mid X_j \geq s\} \end{aligned} \tag{8}$$

where s and j are values that minimize the RSS, given by

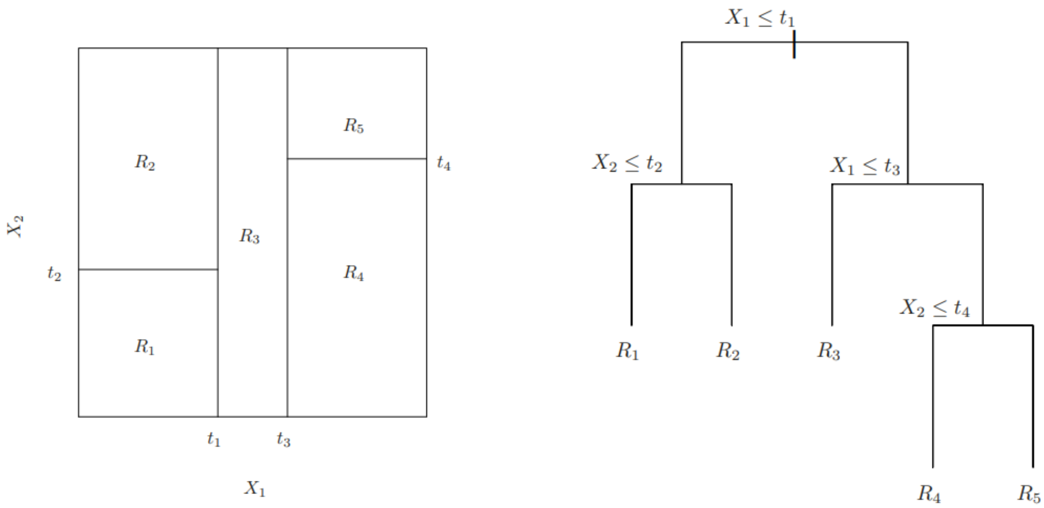
$$\sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \tag{9}$$

where $\hat{y}_{R_1}$ represents the mean response for the training observations in $R_1(j,s)$ and $\hat{y}_{R_2}$ represents the mean response for the training observations in $R_2(j,s)$.

Then, one of the two created regions is divided into another two regions. This process is repeated until a specific stopping criterion is reached, for example, the regions contain a maximum number of observations previously defined. The result is the mean response of the training observations in the region where the test observation is.

In Figure 4, we can see on the left how the process of recursive binary splitting looks like graphically with an example of five partitions of the predictor space. In the same figure, the decision tree that results from this process is presented on the right. The first node is called the root node, and the terminal nodes are called leaves.

Figure 4: Recursive Binary Splitting and Decision Tree



Source: James et al. (2023)

Next, we briefly present the XGBoost. For further details, we refer the reader to Chen & Guestrin (2016).

Let's define a dataset with n examples and m features like $\mathcal{D} = \{(x_i, y_i)\}$ ($|\mathcal{D}| = n$, $x_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$). A predicted value, $\hat{y}_i$, of a tree ensemble model is obtained through K additive functions, as shown in Equation 10.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}, \tag{10}$$

where $\mathcal{F} = \{f(x) = w_{q(x)}\}$ $(q : \mathbb{R}^m \rightarrow T,\ w \in \mathbb{R}^T)$ represents the regression trees space, f_k represents an independent tree structure with T leaves, with each leaf having a weight of w. w_i represents a continuous score on the i-th leaf. For every n, the decision rules in each tree, given by q, determine which leaf it is classified into. Then, the final prediction is obtained by summing all of the scores of each of those leaves, given by w.

To learn these functions, a tree ensemble model minimizes a regularized learning objective given by,

$$\mathcal{L}(\phi) = \sum_i \ell(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{11}$$

$$\text{where} \quad \Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2,$$

- $\ell(\hat{y}_i, y_i)$ defines a differentiable convex loss function that measures the difference between the predicted and the target values, $\hat{y}_i$ and y_i, respectively;
- $\Omega(f)$ is a regularization term that penalizes more, the greater the complexity of the decision trees (given by the regression tree functions). This term helps prevent overfitting.

Besides using regularization to avoid overfitting, shrinkage is also used with this purpose. This way, it is possible to give less importance to individual trees in order for future trees to improve the model.

The XGBoost model can also handle missing data. For each tree node, there is a default direction that is learnt from the data. This means that whenever there is a missing value, that instance is classified taking into account the default direction.

Since, in this case, we only have past parameters to forecast their future values, more features are added to the model, namely lagged parameter values and rolling means of the parameters. This is called feature engineering. According to Bojer (2022), feature engineering aims to transform the dataset in a way that will increase the predictive accuracy of the model. Those transformations can "include external factors, lagged values of the time series, rolling statistics, and other time series features". The author also states that feature selection is often performed using Cross-Validation, the method used in this work and explained next.

The optimal number of lags and the size of the time windows for the rolling means to be added are determined using K-fold Cross-Validation, with K=3. This means the 2024 yield data is split into three equal-sized samples. Two of them are used as the training set, and the other one is used as the testing set. This is repeated three times until each one of the three samples has been used as the testing set once. With this method, different combinations of lagged parameter values and rolling means are tested. The ones with the lower MSE are chosen as the optimal ones. The optimal lag values and rolling mean window sizes are presented in Section 5.3. Regarding the optimal hyperparameter values, those are obtained by performing a Grid Search using K-fold Cross-Validation, with K=3. This process is similar to the one performed for choosing the optimal lags and rolling means. The difference is that the best hyperparameter values are chosen from a pre-specified grid with ranges for the values that the hyperparameters can take.

4.3.3 Constructing the Yield Curve

After obtaining the forecasted parameters with the ARIMA and XGBoost models, they are used as inputs to the equations of the Nelson-Siegel and Svensson models presented in Sections 4.2.1 and 4.2.2. Subsequently, the forecasted January 2025 yield curves are obtained.

4.3.4 Evaluating the Forecast Performance

The forecasted yields are compared with their actual values in January 2025, the testing set. The measures used for evaluating the forecast accuracy are the MSE, the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE), popular performance metrics used for forecasting purposes, referred in Botchkarev (2018). The lower the values, the better is the quality of the forecast, since the forecasted values are closer to the observed values. The values of the RMSE and the MAE are expressed in percentage points.

The RMSE is the squared root of the MSE, explained in Equation (3).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Actual}_i - \text{Forecasted}_i)^2} \tag{12}$$

The MAE is the mean absolute difference between the actual and the forecasted yields.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\text{Actual}_i - \text{Forecasted}_i| \tag{13}$$

The MAPE measures the mean absolute proportion of the deviation between the actual and the predicted yields in the value of the actual yield.

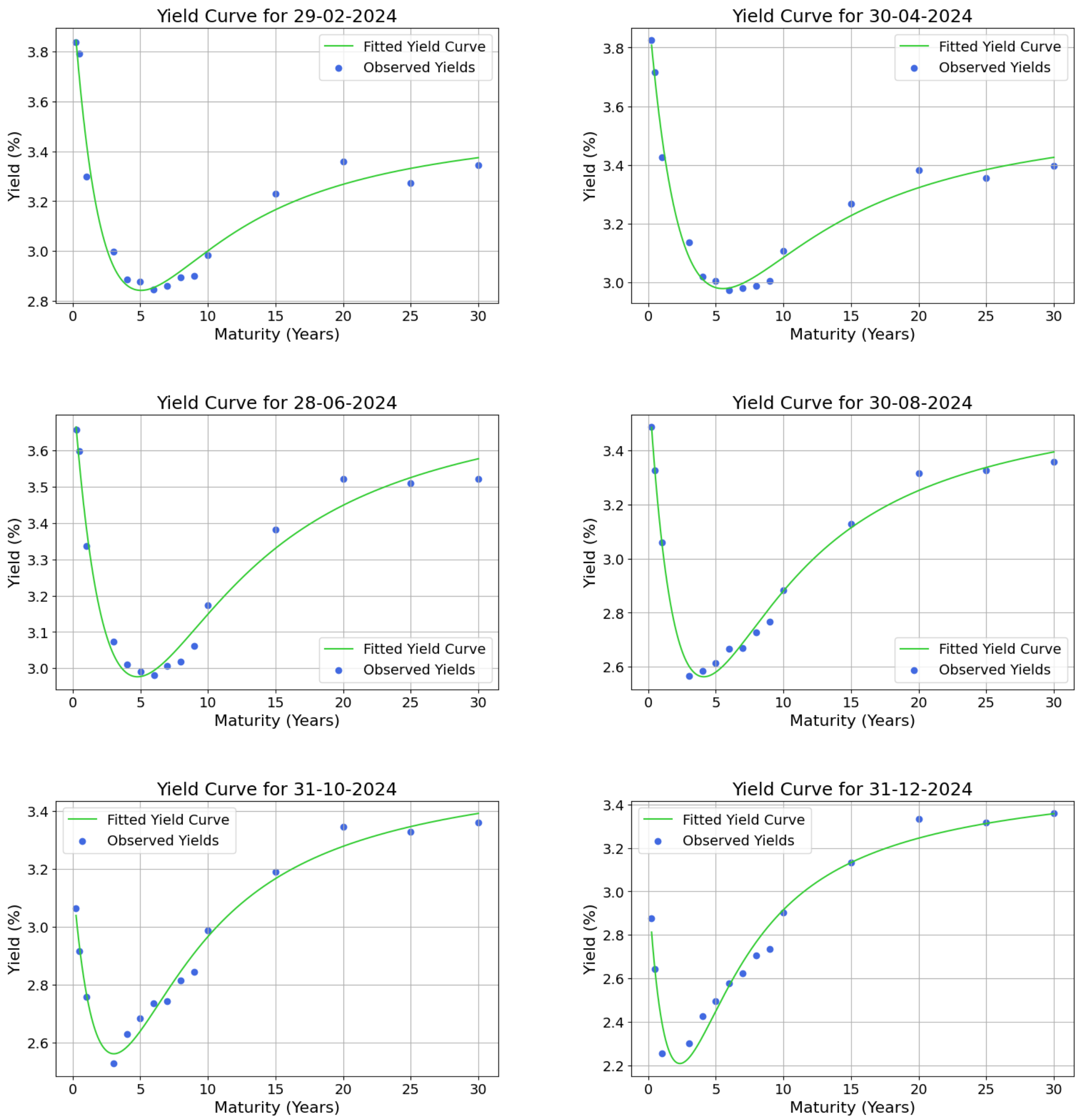
$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\text{Actual}_i - \text{Forecasted}_i}{\text{Actual}_i}\right| \tag{14}$$

5 RESULTS

5.1 2024 Yield Curves

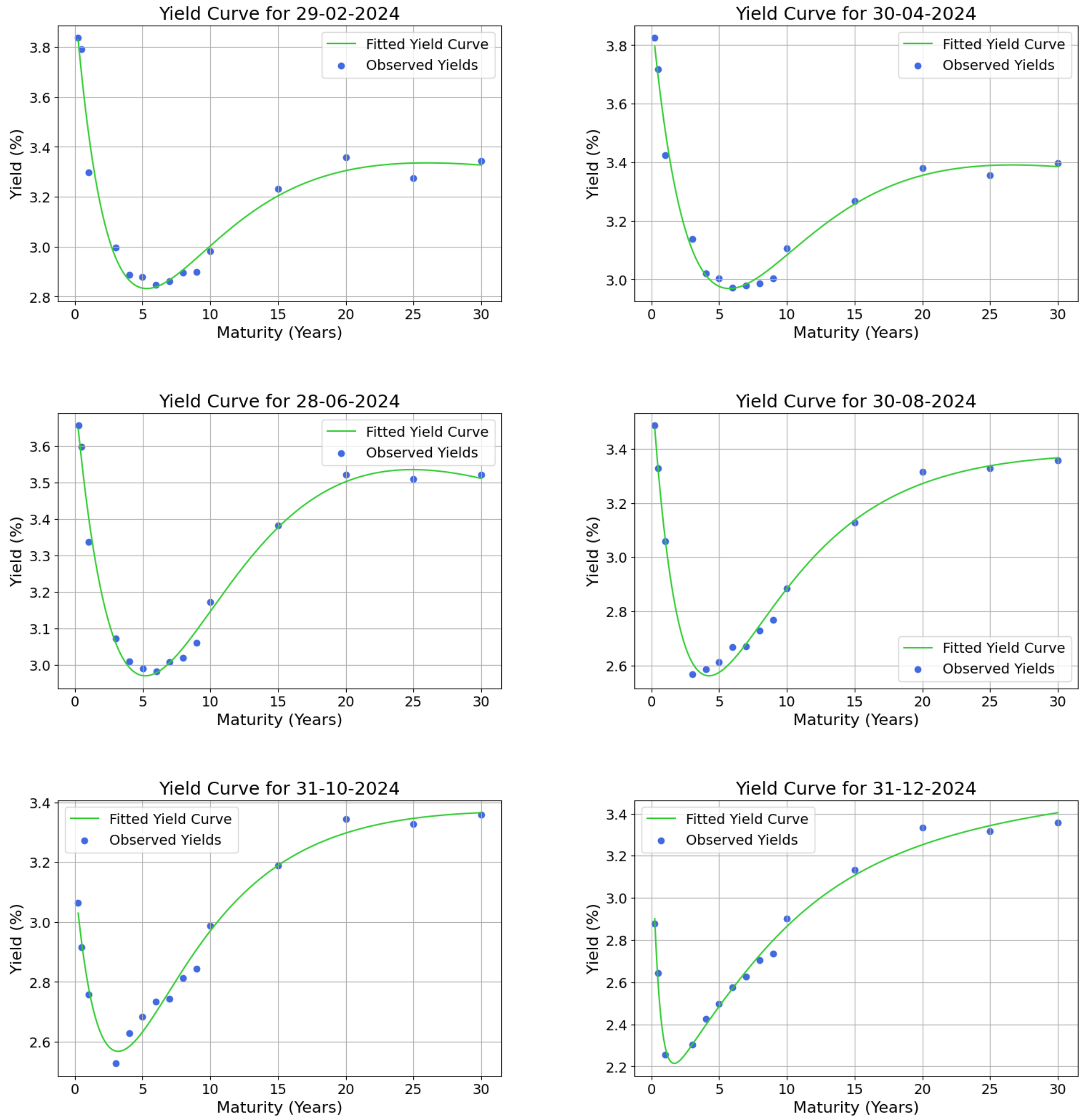
The yield curves presented in Figures 5 and 6 are obtained using the fitted parameters from Equations (1) and (2), respectively. The yield curves presented refer to six dates during 2024.

Figure 5: Nelson-Siegel Model 2024 Yield Curves



2024 yield curves calibrated with the Nelson-Siegel model for selected end-of-month dates in February, April, June, August, October and December.

Figure 6: Svensson Model 2024 Yield Curves



2024 yield curves calibrated with the Svensson model for selected end-of-month dates in February, April, June, August, October and December.

In Table II, we can see the daily average of the MSE and the Adjusted R^2, denoted as R^2_{adj}, registered in 2024, for each model.

Table II: Average MSE and R^2_{adj} of the Calibrated Yield Curves

	Nelson-Siegel	Svensson
MSE	0.0025	0.0018
R^2_{adj}	0.9666	0.9708

It can be confirmed that both models represent a good fit for the data, presenting low MSE values and Adjusted R^2 values close to 1. However, the Svensson model presents slightly better results. This is expected, because the Svensson model adds two parameters, β_3 and τ_2, to the Nelson-Siegel model, increasing its complexity but also its flexibility. Consequently, the Svensson model adapts better to the data.

In Figures A.1 and A.2, we can see the residuals plot per maturity for both models. The absolute value of the residuals decreased in the first months of the year and remained stable throughout the second half of the year, except for the residuals of the 1 year and 9 year maturity points in the Nelson-Siegel model, which showed an increase, in absolute terms, during the last two months of the year.

5.2 *Fitted Parameters*

After fitting both models to the data, the time series of parameters are obtained, comprising the period from January 2024 until December 2024. They are portrayed in Figures 7 and 8.

In Tables III and IV, the descriptive statistics of the parameters are presented.

Figure 7: Nelson-Siegel Model Parameters

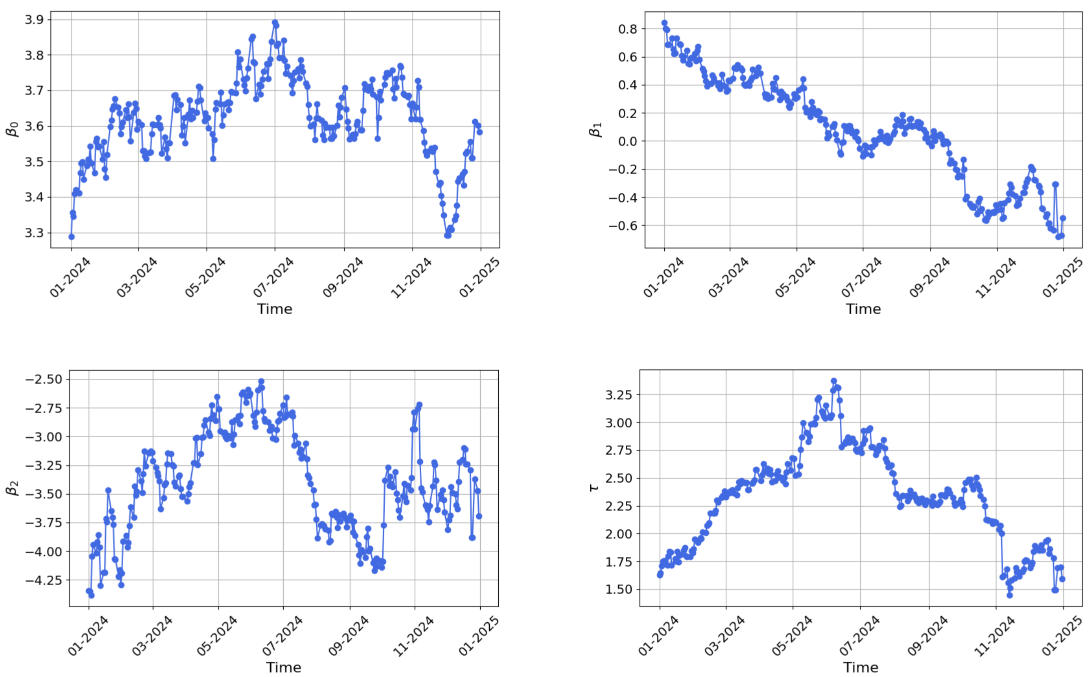


Table III: Descriptive Statistics of the Nelson-Siegel Model Parameters

Parameter	Mean	Std. dev.	Minimum	Q1	Q2 (Median)	Q3	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(7)$	$\hat{\rho}(30)$
β_0	3.6179	0.1170	3.2876	3.5531	3.6226	3.6993	3.8930	0.9418	0.6939	−0.0271
β_1	0.0739	0.3684	−0.6826	−0.2021	0.0819	0.3839	0.8446	0.9875	0.9490	0.8224
β_2	−3.4043	0.4390	−4.3849	−3.7297	−3.4271	−3.0151	−2.5150	0.9518	0.7191	0.3053
τ	2.3295	0.4324	1.4448	1.9527	2.3424	2.5814	3.3776	0.9869	0.9072	0.4756

Using the interpretation of the parameters explained in Chapter 4, we can see that for the Nelson-Siegel model, on average, the curve level is 3.62%, the slope is 0.07%, and the curvature is -3.40.

Figure 8: Svensson Model Parameters

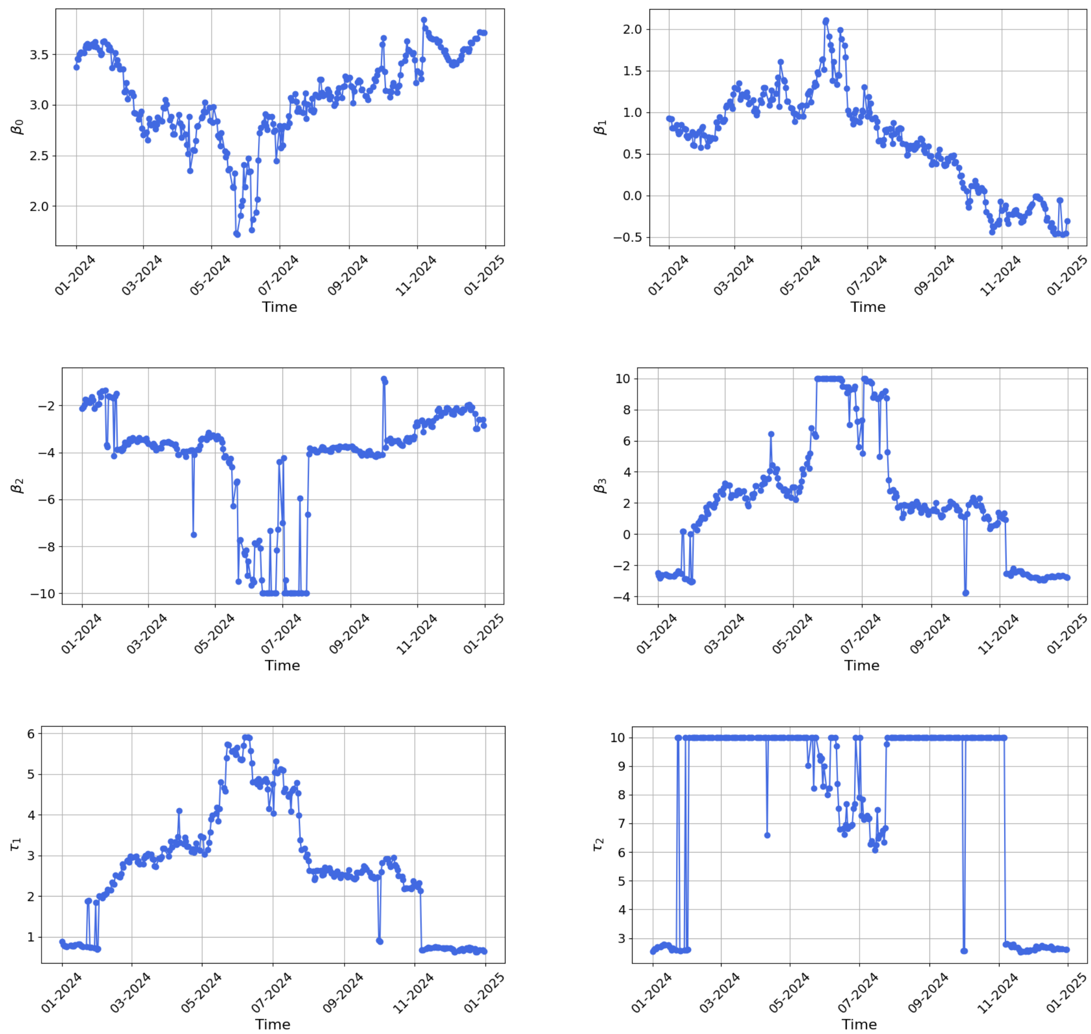


Table IV: Descriptive Statistics of the Svensson Model Parameters

Parameter	Mean	Std. dev.	Minimum	Q1	Q2 (Median)	Q3	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(7)$	$\hat{\rho}(30)$
β_0	3.0793	0.4280	1.7161	2.8218	3.1102	3.4459	3.8460	0.9577	0.8061	0.4416
β_1	0.6520	0.5948	-0.4745	0.1203	0.7395	1.0809	2.1097	0.9779	0.9065	0.7430
β_2	-4.2902	2.3660	-10.0000	-4.0879	-3.7224	-2.9842	-0.8419	0.9184	0.7742	0.2779
β_3	2.2819	3.9068	-3.7861	0.3174	1.8646	3.4441	10.0000	0.9685	0.8731	0.4282
τ_1	2.7147	1.4519	0.6200	1.9781	2.6448	3.3214	5.9127	0.9781	0.9017	0.4651
τ_2	7.8690	3.0627	2.5063	6.3659	10.0000	10.0000	10.0000	0.8921	0.7092	0.0060

For the Svensson model, on average, the curve level is 3.08%, the slope is 0.65%, the first curvature is -4.29, and the second curvature is 2.28.

Before proceeding to the forecast, and to determine if these time series are stationary or non-stationary, the ADF test is performed, with a degree of confidence of 95%. In this test, the null hypothesis states that the series is non-stationary and the alternative hypothesis states that the series is stationary.

The results of the test are shown in Table V.

Table V: ADF Test Results

Parameter	Nelson-Siegel		Svensson	
	Test Statistic	p-value	Test Statistic	p-value
β_0	-3.5153	0.0076	-1.4622	0.5521
β_1	-1.1634	0.6891	-0.8549	0.8025
β_2	-2.7425	0.0670	-2.4686	0.1233
β_3			-1.5342	0.5165
τ_1	-1.1170	0.7082	-1.2015	0.6729
τ_2			-1.8402	0.3607

If the p-value is lower than the significance level, we should reject the null hypothesis and conclude the series are stationary. If the p-value is higher than the significance level, we should not reject the null hypothesis and conclude the series are non-stationary.

For all parameters, except for β_0 in the Nelson-Siegel model, the p-value is higher than 0.05 (significance level). This indicates that β_0 in the Nelson-Siegel model is stationary and the rest of the parameters are non-stationary, meaning they need to be differentiated in order to be analysed and forecasted.

However, when plotting the Autocorrelation Function (ACF) for the parameters of the

Nelson-Siegel model (see Figure A.3), we can observe a gradual decay of the values of this function for β_0. This is an indicator that the parameter is not, in fact, stationary and needs to be differentiated as well.

So, for forecasting purposes, ARIMA is a good model because it incorporates the differencing of non-stationary variables. The degree of differencing, presented in Section 5.3, is determined by the package *pmdarima* in Python, as referred in Chapter 4.

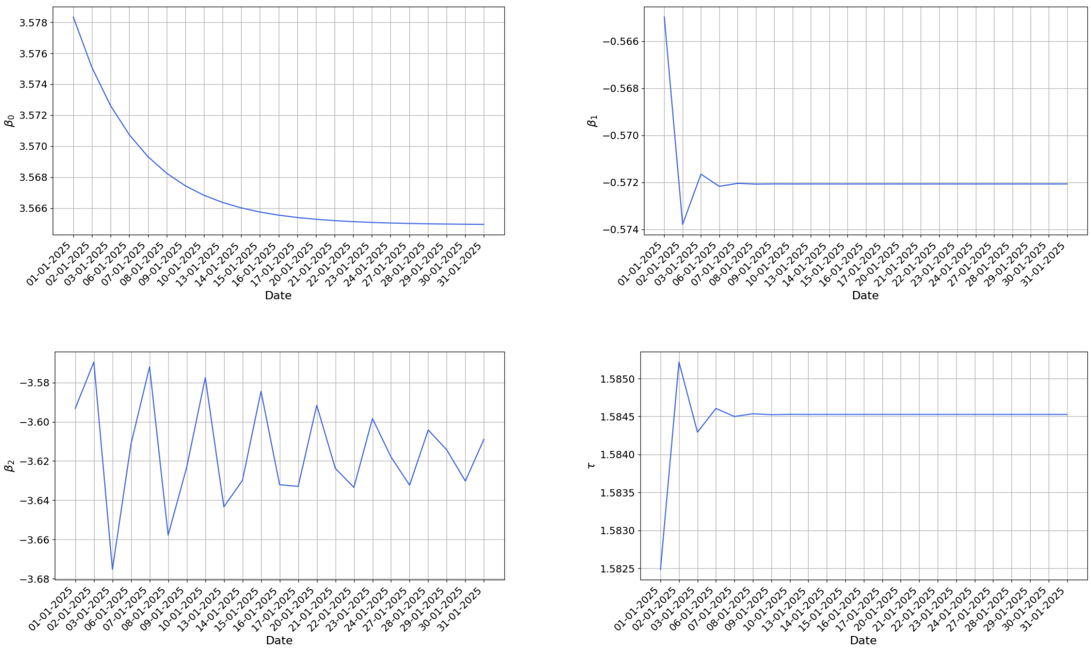
The XGBoost also predicts the same variables as the model above. In this case, it not only uses the past values of the parameters to predict their future values like the previous method, but it also incorporates variables like the lagged parameter values and the rolling means of the parameters. Their values are specified in Section 5.3.

5.3 Forecasted Parameters

Using the ARIMA and XGBoost methods, in Figures 9 and 10, we can see the plot of the forecasted parameters for the Nelson-Siegel model, from the 1st until the 31st of January 2025.

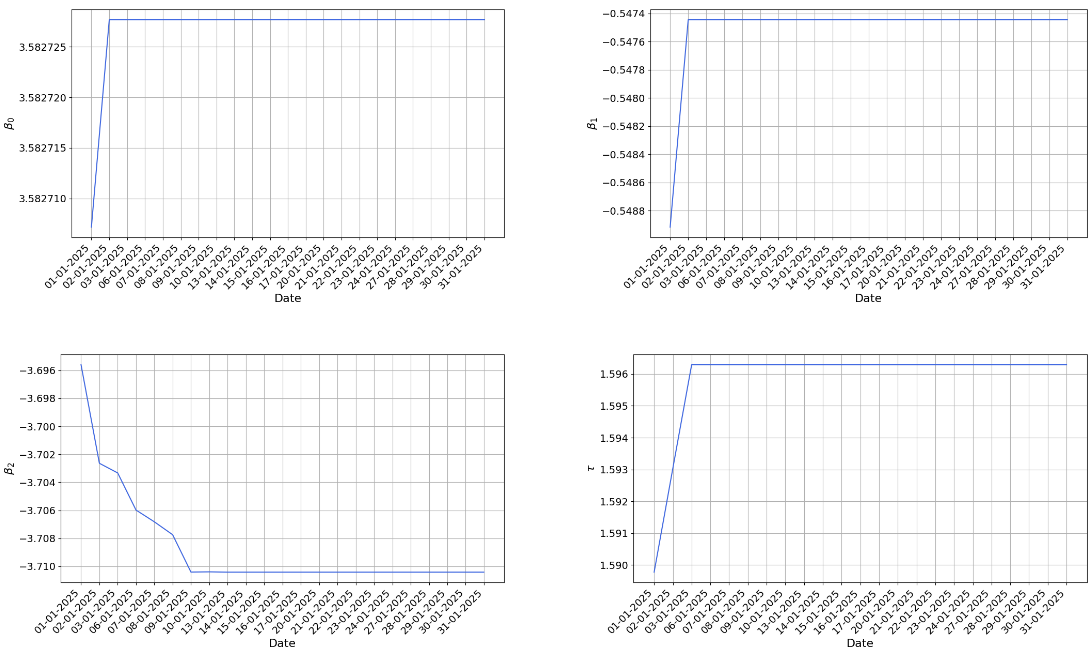
Regarding Figure 9, the optimal orders found for each parameter are: ARIMA(1,1,1) for β_0, ARIMA(1,1,2) for β_1, ARIMA(2,1,2) for β_2, and ARIMA(1,1,1) for τ.

Figure 9: Nelson-Siegel Model Forecasted Parameters using ARIMA



Besides the past parameters values, the forecast results in Figure 10 are estimated by calibrating the XGBoost model also using the parameters values with an optimal 1 day lag for β_0, β_1 and τ. For β_2, a lag of 2 days is considered. A 1 day optimal rolling mean is a feature used for all the parameters.

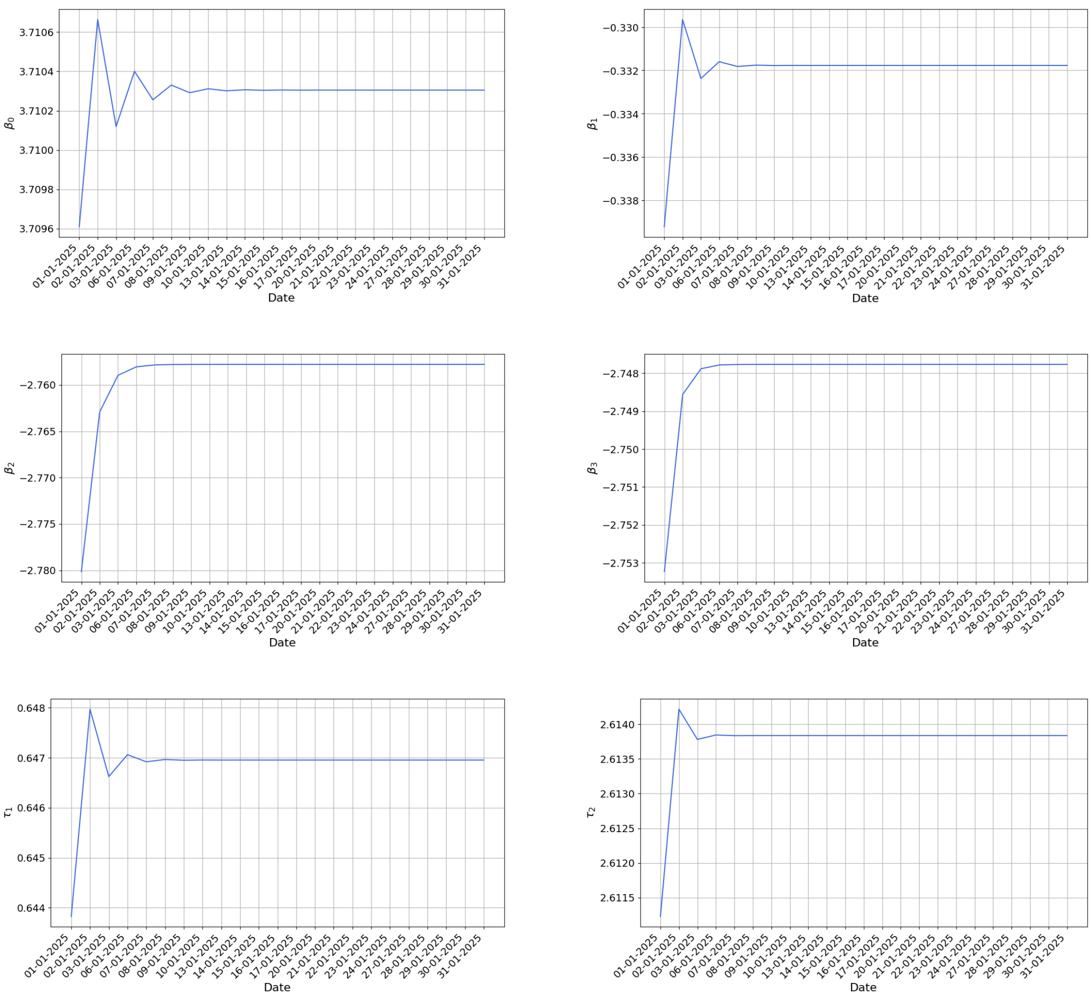
Figure 10: Nelson-Siegel Model Forecasted Parameters using XGBoost



For the Svensson model, the forecasted parameters, from the 1st until the 31st of January 2025, are displayed in Figures 11 and 12, using the ARIMA and XGBoost models, respectively.

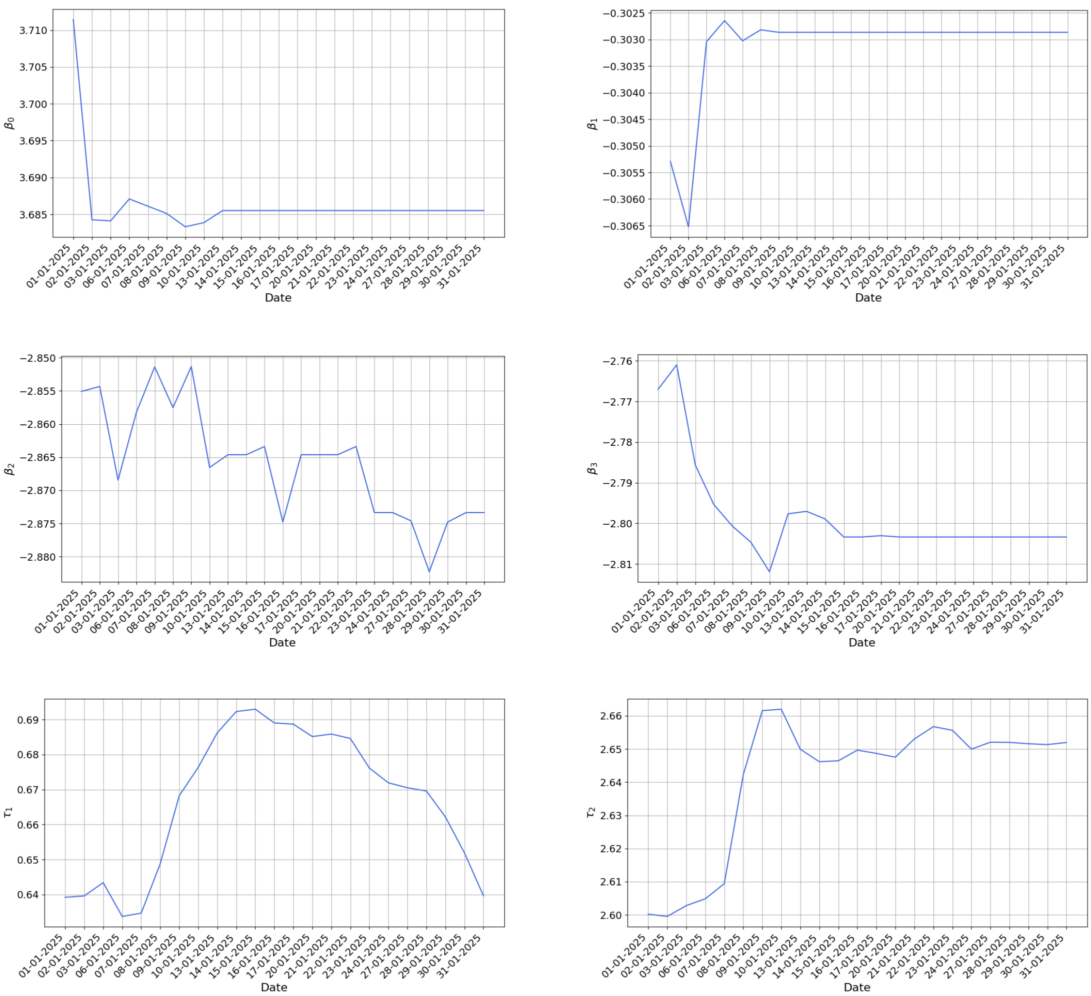
The forecasted parameters displayed in Figure 11 are obtained with the following optimal ARIMA models: ARIMA(1,1,1) for β_0, ARIMA(1,1,1) for β_1, ARIMA(1,1,1) for β_2, ARIMA(1,1,2) for β_3, ARIMA(1,1,2) for τ_1, and ARIMA(1,1,2) for τ_2.

Figure 11: Svensson Model Forecasted Parameters using ARIMA



Regarding Figure 12, the optimal number of lags considered in the lagged features during the estimation of XGBoost are different than the ones used for the Nelson-Siegel model. In this case, apart from the past parameters values, the lags considered for the parameters lagged values are: 8 days for β_0 and β_2, 3 days for β_1, 10 days for β_3, and 9 days for τ_1 and τ_2. A 1 day optimal rolling mean feature is used for all the parameters, as in the previous model.

Figure 12: Svensson Model Forecasted Parameters using XGBoost



Throughout all of the forecasts, it is observable that the parameters do not vary a lot in value when compared to the fitted parameters, and are, frequently, constant. This can be explained by the fact that the fitted parameters, used to train the forecasting models, present a low standard deviation during 2024 (see Tables III and IV), impacting the forecasting ability of the models.

5.4 January 2025 Forecasted Yield Curves

This Section presents the forecasted yield curves for 1 day, 1 week (7$^{\text{th}}$ of January) and 1 month (31$^{\text{st}}$ of January) in January 2025, as well as the metrics that evaluate the forecast performance of the models used: MSE, RMSE, MAE, and MAPE.

The yield curves presented in Figures 13, 14, 15, and 16 are obtained using as input the predicted parameters presented in Section 5.3 in Equations (1) and (2).

Figure 13: Nelson-Siegel Model/ARIMA Forecasted Yield Curves

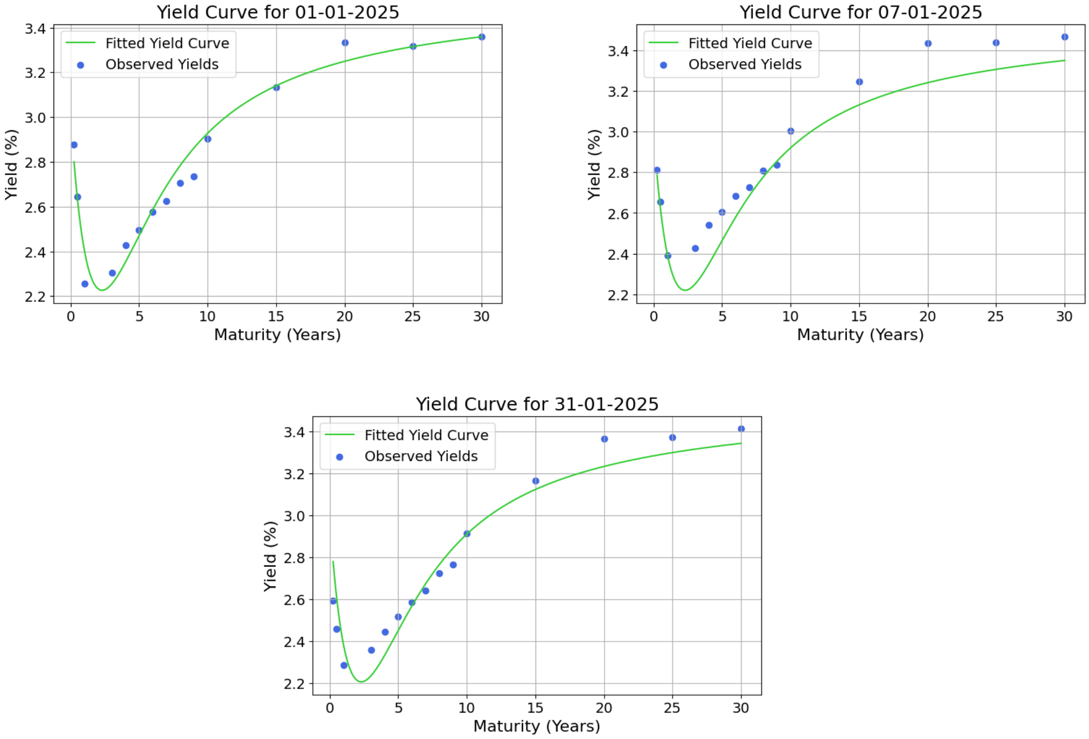


Figure 14: Nelson-Siegel Model/XGBoost Forecasted Yield Curves

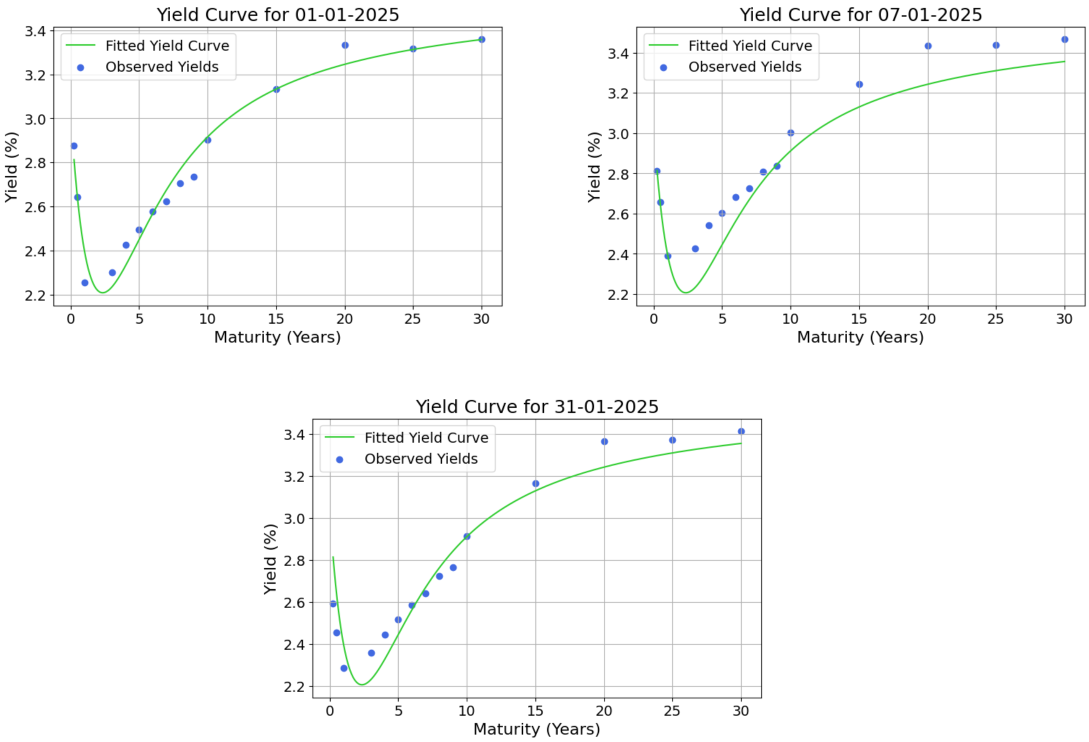


Figure 15: Svensson Model/ARIMA Forecasted Yield Curves

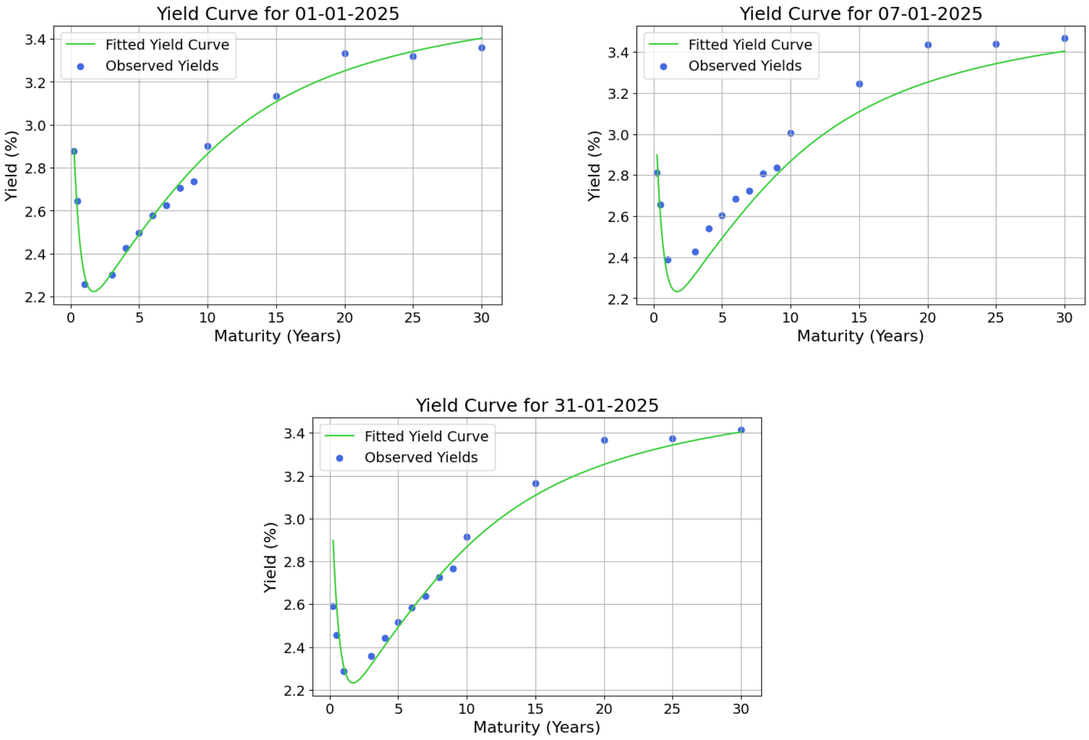
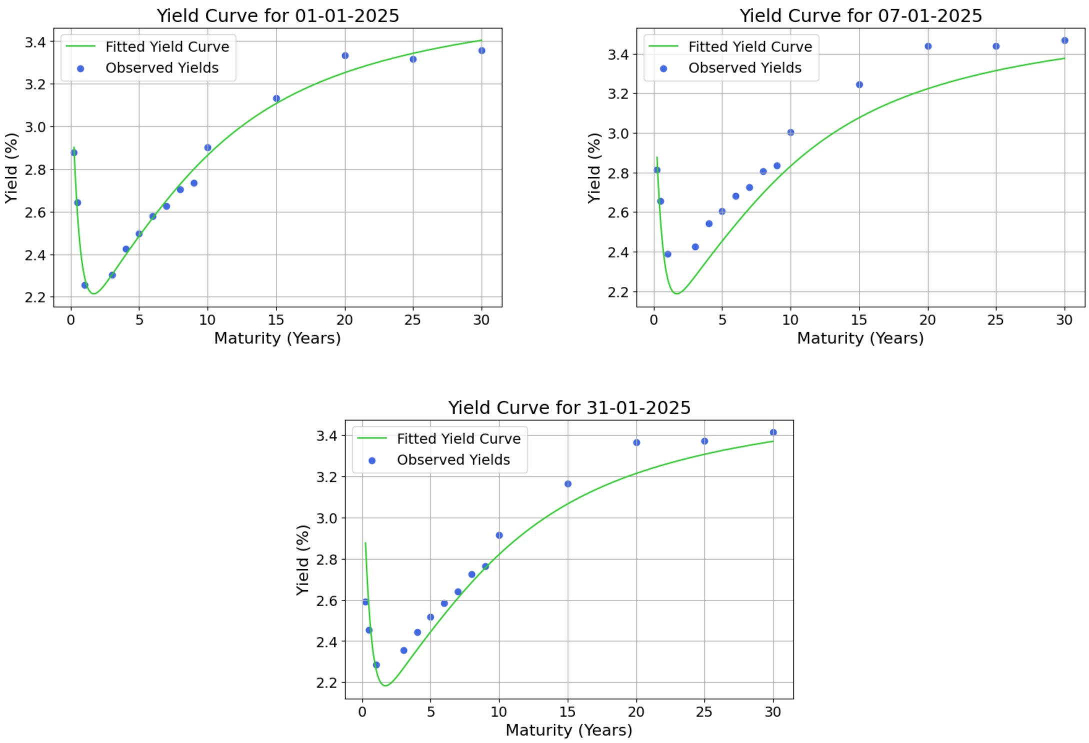


Figure 16: Svensson Model/XGBoost Forecasted Yield Curves



In Tables VI, VII, VIII, and IX are shown the values of MSE, RMSE, MAE, and MAPE, respectively, for three different forecast horizons. For comparability, the values portrayed for the last two forecast horizons are the daily averages of those metrics. RMSE and MAE values are expressed in percentage points.

Table VI: Average MSE for Each Forecasting Model

Forecast Horizon	Nelson-Siegel		Svensson	
	ARIMA	XGBoost	ARIMA	XGBoost
1 day	0.0048	0.0046	0.0016	0.0015
1 week	0.0087	0.0090	0.0060	0.0104
1 month	0.0228	0.0232	0.0200	0.0339

Table VII: Average RMSE for Each Forecasting Model

Forecast Horizon	Nelson-Siegel		Svensson	
	ARIMA	XGBoost	ARIMA	XGBoost
1 day	0.0690	0.0677	0.0394	0.0391
1 week	0.0911	0.0927	0.0732	0.0952
1 month	0.1446	0.1463	0.1338	0.1747

Table VIII: Average MAE for Each Forecasting Model

Forecast Horizon	Nelson-Siegel		Svensson	
	ARIMA	XGBoost	ARIMA	XGBoost
1 day	0.0525	0.0513	0.0325	0.0332
1 week	0.0741	0.0751	0.0641	0.0882
1 month	0.1275	0.1293	0.1225	0.1648

Table IX: Average MAPE for Each Forecasting Model

Forecast Horizon	Nelson-Siegel		Svensson	
	ARIMA	XGBoost	ARIMA	XGBoost
1 day	0.0201	0.0198	0.0115	0.0118
1 week	0.0266	0.0272	0.0226	0.0313
1 month	0.0445	0.0455	0.0428	0.0577

The error values increase with the dimension of the forecasting horizon. This is expected because we are further away from the end point of the data used to train the models and obtain the values of the forecasted parameters.

Firstly, for the 1 day forecast horizon, the MSE and RMSE metrics agree that using the forecasted parameters obtained with the XGBoost model as input for the Svensson model yields the best results, obtaining a 0.0015 MSE and 0.0391 p.p. RMSE. However, if we look at MAE and MAPE, it suggests that using the forecasted parameters of the ARIMA model combined with the Svensson model is better, showing the smallest values of MAE and MAPE, 0.0325 p.p. and 1.15%, respectively. This means that the forecasted yields have, on average, a 3.25 b.p. difference from the actual yields, in absolute terms. The worst results for this time horizon are obtained using the predicted parameters of ARIMA as input for the Nelson-Siegel model.

The MSE and the MAE are similar, but they have an important difference. The MSE is more sensitive to extreme values because it amplifies the difference between the actual and forecasted values by squaring that difference. That does not happen in MAE, which considers only the module of the difference. So, the MSE and, consequently, the RMSE penalises more these extreme values, showing higher errors. This needs to be taken into account when choosing the best model. So, the XGBoost model used in conjunction with the Svensson model predicts smoother values than ARIMA and the Svensson model used together, that show higher MSE and RMSE values, but lower MAE and MAPE values.

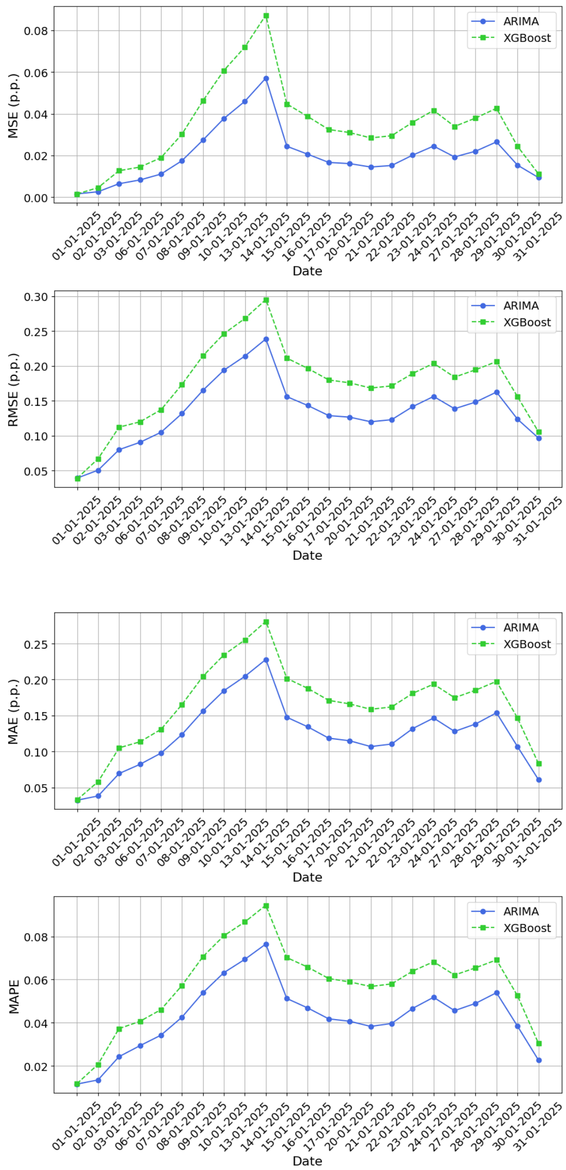
Secondly, for the 1 week forecast horizon, all the measures point to the same best models: ARIMA forecasted parameters used in the Svensson model. Using these models combined, the MSE is 0.0060, the RMSE is 0.0732 p.p., the MAE is 0.0641 p.p., and the MAPE is 2.26%. So, the forecasted yields have, on average, a 6.41 b.p. deviation from the real values, in absolute terms. In contrast to the 1 day time horizon, for the 1 week ahead forecast, the worst results are achieved when using the XGBoost and the Svensson models together.

Lastly, for the 1 month forecast horizon, the models that give the best results are in line with the results obtained for the 1 week forecast. The forecasted parameters obtained with ARIMA used as inputs for the Svensson model provide the best results. The MSE is 0.02, the RMSE is 0.1338 p.p., the MAE is 0.1225 p.p. and the MAPE is 4.28%. So, when forecasting for 1 month, the yields show, on average, a 12.25 b.p. error compared to the observed values, in absolute terms. The worst results are again given by the XGBoost model used with the Svensson model.

In Figures 17 and 18, the plots for each of the metrics are presented for the Nelson-Siegel and Svensson models, respectively, highlighting the differences between the two forecasting models used, ARIMA and XGBoost.

Figure 17: MSE, RMSE, MAE and MAPE: Nelson-Siegel Model

Figure 18: MSE, RMSE, MAE and MAPE: Svensson Model



In Figure 17, we can verify that the results obtained with ARIMA and XGBoost models are very similar, in terms of the error metrics chosen. However, when using the Svensson model to derive the yield curve, the difference in results between the forecasting models is more striking. The XGBoost combined with the Svensson model creates larger overall errors than any other combination of models, as shown previously by analyzing the 1 week and 1 month forecasts.

6 Conclusion

In this research, we applied two fitting models, the Nelson-Siegel and the Svensson, and two forecasting models, the ARIMA and the XGBoost, to calibrate and predict the yield curve of EU-Bonds.

The XGBoost model is better at forecasting for smaller windows of time, such as 1 day. As the time window expands, its forecast accuracy decreases. Contrasting with this model, ARIMA shows better results for longer forecasting horizons.

Regarding the comparison between the Nelson-Siegel and the Svensson models, we can observe that for the 1 week and 1 month forecast horizons, the Nelson-Siegel model presents intermediary results. This model is expected to present the worst results for every scenario because it is a simpler model that has fewer explanatory parameters of the yield curve. However, that only happens for the 1 day forecasting window.

Despite these results, it is important to highlight that all models used in this work yield good results for forecasting purposes. The highest MSE value registered is 0.0339, the highest RMSE is 0.1747 p.p., the highest MAE is 0.1648 p.p., and the highest MAPE is 5.77%, when considering a 1 month forecasting horizon and a combination of the Svensson and XGBoost models. This means that, at most, the error between the real yield value and the forecasted yield value is, on average, 16.48 b.p., in absolute terms. Although these are the worst results, they are still good, considering the corresponding forecasting window.

Our research outperforms several studies presented in Chapter 3. In our work, for the 1 month forecast horizon, the worst RMSE and MAE values obtained are 0.1747 p.p. and 0.1648 p.p., respectively, when considering a combination of the Svensson and XGBoost models. In Diebold & Li (2006), the lowest RMSE obtained was 0.2458 p.p., when using a random walk model. In Christensen et al. (2007), the lowest RMSE attained was 0.2833 p.p., considering an arbitrage-free dynamic Nelson-Siegel model with independent factors. In Lee (2023), the lowest RMSE and MAE values achieved were 0.2559 p.p. and 0.1872 p.p., respectively, with a 4-factor deep learning Nelson-Siegel model using a recurrent neural network. All these results correspond to the daily averages of each metric, considering an out-of-sample 1 month ahead forecast.

There are, of course, some limitations in this work that can have an impact on our results.

Firstly, the data pool is limited because the EU-Bonds are a relatively recent product, still evolving. So, there is not daily yield data available for previous years for all the maturities selected for this study.

Secondly, not having a lot of data can hamper the results, especially for the ML model used, XGBoost. According to Cerqueira et al. (2022), the forecasting power of ML models increases as they are trained with more data.

Thirdly, for predicting future yield values we are only using past values of the same variable. This imposes a limitation since other factors can explain changes in the yield curve. In the article by Koroleva & Kopeykin (2022), the authors conclude that macroeconomic factors such as the price of gold and oil, inflation and Gross Domestic Product (GDP) per capita influence the yields of government bonds.

Various extensions of this work could be done. For example, different models containing other explanatory variables of the yield values of EU-Bonds could be used to forecast the yield curve.

It could also be developed a hybrid model that combines a statistical model with a ML model, capturing, respectively, the linear and non-linear factors affecting the yield curve.

This research can also be extended to other products. Yield forecasts for different products can be useful, as well as the spreads between them.

It can also be interesting to study how the EU-Bonds yield curve relates to other EU government bonds yield curves. It could be investigated if the movement in one yield curve has an impact on the other, and if so, what would be the magnitude of that impact.

REFERENCES

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

BIS (2005), Zero-coupon yield curves: Technical documentation, Technical Report 25, Bank for International Settlements.
URL: *https://www.bis.org/publ/bppdf/bispap25.pdf*

BIS (2021), Big data and machine learning in central banking, BIS Working Papers 930, Bank for International Settlements.
URL: *https://www.bis.org/publ/work930.pdf*

Bojer, C. S. (2022), 'Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities', *International Journal of Forecasting* **38**(4), 1555–1561. Special Issue: M5 competition.
URL: *https://www.sciencedirect.com/science/article/pii/S0169207021001771*

Botchkarev, A. (2018), 'Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology', *arXiv preprint arXiv:1809.03006* .

Box, G. E. P. & Jenkins, G. M. (1970), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2016), *Time Series Analysis: Forecasting and Control*, 5th edn, John Wiley & Sons, Inc., Hoboken, New Jersey.

Castellani, M. & dos Santos, E. A. (2006), Forecasting long-term government bond yields: An application of statistical and ai models, Working Paper WP 04/2006/DE, ISEG – Lisbon School of Economics and Management, Department of Economics, Universidade de Lisboa.
URL: *https://depeco.iseg.ulisboa.pt/wp/wp042006.pdf*

Castello, O. & Resta, M. (2024), 'Optimal time varying parameters in yield curve modeling and forecasting: A simulation study on brics countries', *Comput. Econ.* **65**(4), 2081–2113.
URL: *https://doi.org/10.1007/s10614-024-10619-z*

Cerqueira, V., Torgo, L. & Soares, C. (2022), 'A case study comparing machine learning with statistical methods for time series forecasting: size matters', *Journal of Intelligent Information Systems* **59**.

Chen, T. & Guestrin, C. (2016), Xgboost: A scalable tree boosting system, *in* 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '16, ACM, p. 785–794.
URL: *http://dx.doi.org/10.1145/2939672.2939785*

Christensen, J. H. E., Diebold, F. X. & Rudebusch, G. D. (2007), The affine arbitrage-free class of: Nelson-siegel term structure models, Working Paper 13611, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w13611*

Christensen, J. H. E., Diebold, F. X. & Rudebusch, G. D. (2009), 'An arbitrage-free generalized nelson—siegel term structure model', *The Econometrics Journal* **12**(3), C33–C64.
URL: *http://www.jstor.org/stable/23116046*

Cox, J., Ingersoll, J. & Ross, S. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**(2), 385–407.

Diebold, F. X. & Li, C. (2006), 'Forecasting the term structure of government bond yields', *Journal of Econometrics* **130**(2), 337–364.
URL: *https://www.sciencedirect.com/science/article/pii/S0304407605000795*

ECB (2023), Euro area yield curves – technical notes, Technical report, European Central Bank.
URL: *https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/shared/pdf/technical_notes.pdf*

Esaki, T. (2021), 'Appropriate evaluation measurements for regression models', *Chem-Bio Informatics Journal* **21**, 59–69.

Eurex (2025), 'Euro–EU bond futures'.
URL: *https://www.eurex.com/resource/blob/4466676/2d1160e49a724aeafa3e4104755a9921/data/factsheet-euro-EU-bond-futures.pdf*

European Parliament, Policy Department for Budgetary Affairs (2023), 'The rising cost of european union borrowing and what to do about it'.
URL: *https://www.europarl.europa.eu/RegData/etudes/IDAN/2023/749450/IPOL_IDA(2023)749450_EN.pdf*

Fisher, M., Nychka, D. & Zervos, D. (1995), Fitting the term structure of interest rates with smoothing splines, Finance and Economics Discussion Series 95-1, Board of Gov-

ernors of the Federal Reserve System (U.S.).
URL: *https://EconPapers.repec.org/RePEc:fip:fedgfe:95-1*

James, G., Witten, D., Hastie, T., Tibshirani, R. & Taylor, J. (2023), *An Introduction to Statistical Learning with Applications in Python*, Springer Texts in Statistics, Springer, Cham.
URL: *https://link.springer.com/book/10.1007/978-3-031-38747-0*

Jeaab, K., Saoudi, Y. & Falloul, M. E. M. (2024), 'A comparison of lstm, gru, and xgboost for forecasting morocco's yield curve', *Mathematical Modeling and Computing* **11**(3), 674–681.

Koroleva, E. & Kopeykin, M. (2022), 'Understanding of macro factors that affect yield of government bonds', *Risks* **10**(8).
URL: *https://www.mdpi.com/2227-9091/10/8/166*

Lee, S.-H. (2023), 'Yield curve forecasting using deep learning nelson-siegel model'.

McCulloch, J. H. (1975), 'The tax-adjusted yield curve', *Journal of Finance* **30**(3), 811–30.
URL: *https://EconPapers.repec.org/RePEc:bla:jfinan:v:30:y:1975:i:3:p:811-30*

Nelson, C. R. & Siegel, A. F. (1987), 'Parsimonious modeling of yield curves', *The Journal of Business* **60**(4), 473–489.

Rezende, R. B. & Ferreira, M. S. (2008), Modeling and forecasting the brazilian term structure of interest rates by an extended nelson-siegel class of models: A quantile autoregression approach, Research report, Escola Brasileira de Economia e Finanças (FGV EPGE).
URL: *https://epge.fgv.br/finrio/myreview/FILES/CR2/p51.pdf*

Sambasivan, R. & Das, S. (2017), 'A statistical machine learning approach to yield curve forecasting'.
URL: *https://arxiv.org/abs/1703.01536*

Svensson, L. E. O. (1994), 'Estimating and interpreting forward interest rates: Sweden 1992-1994'.
URL: *https://www.elibrary.imf.org/view/journals/001/1994/114/001.1994.issue-114-en.xml*

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics* **5**(2), 177–188.
URL: *https://www.sciencedirect.com/science/article/pii/0304405X77900162*

Zhang, J., Chen, Y., Klotz, S. & Lim, K. G. (2017), ‘International yield curve prediction with common functional principal component analysis’, *Robustness in econometrics* pp. 287–304.

Appendix

Figure A.1: Time Series of the Nelson-Siegel Model Residuals per Maturity (January 2024 – December 2024)

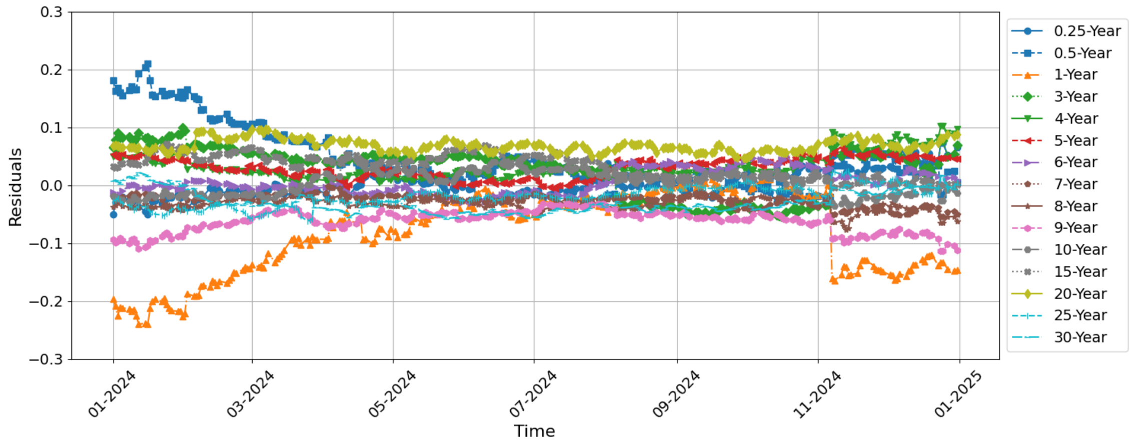


Figure A.2: Time Series of the Svensson Model Residuals per Maturity (January 2024 – December 2024)

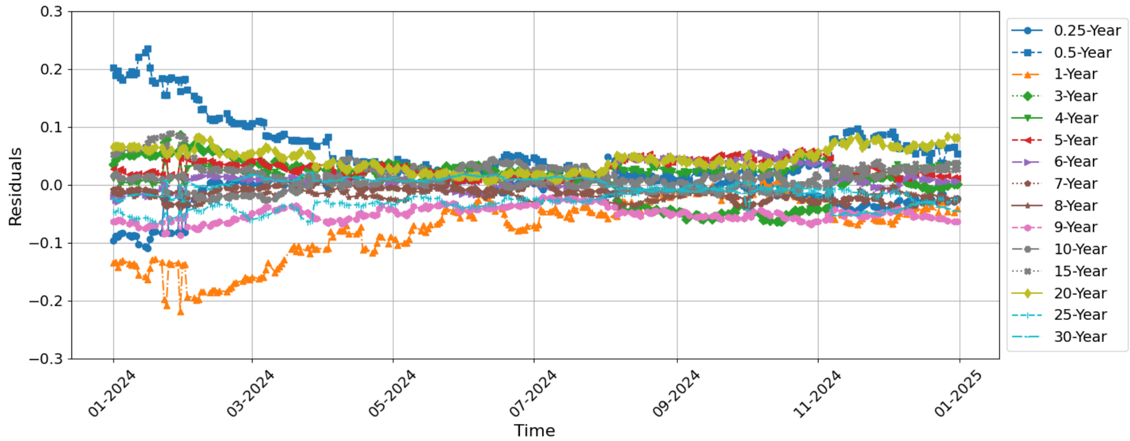
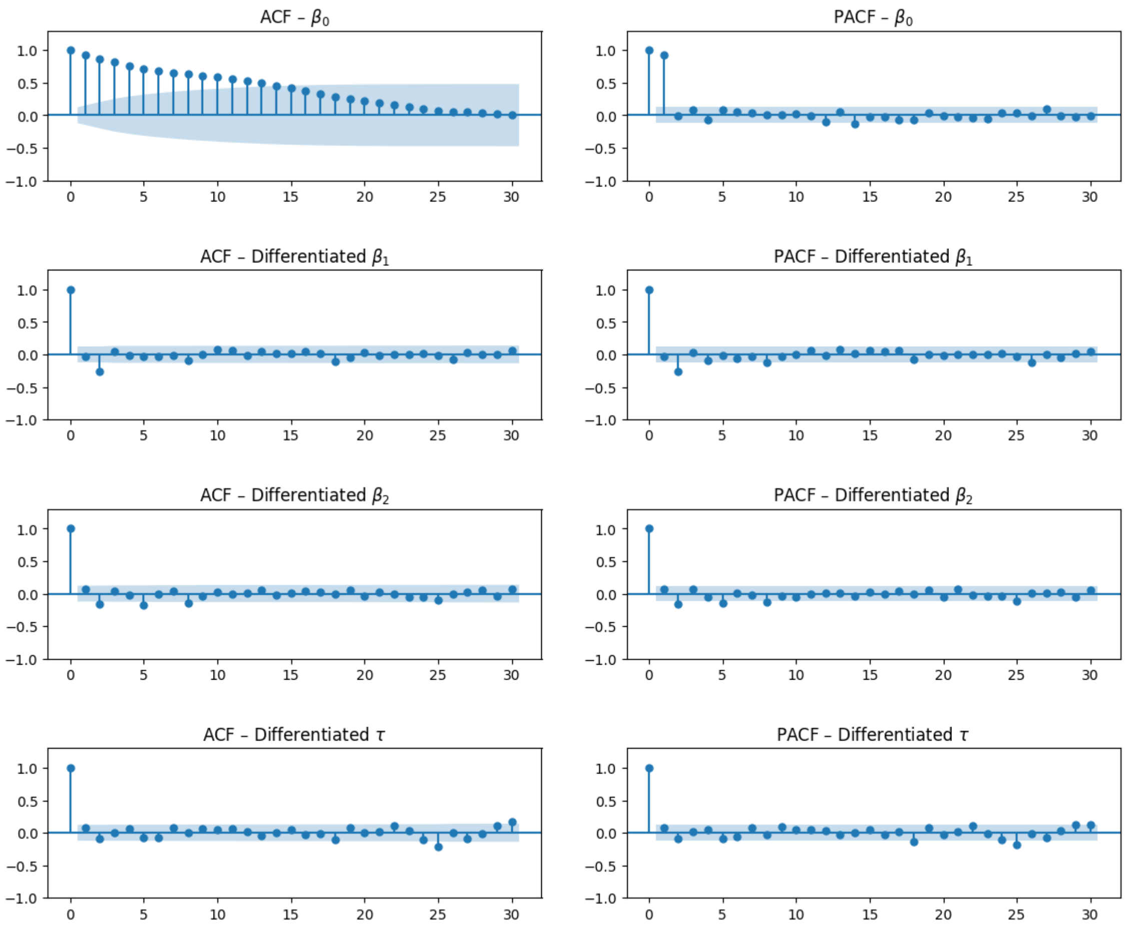


Figure A.3: Autocorrelation and Partial Autocorrelation Functions of the Nelson-Siegel Model Parameters



Disclaimer

In order to be fully transparent, I am writing this disclaimer to explain how AI was employed in this Master's Final Work.

AI tools were used to assist with technical matters, regarding programming issues and formatting errors experienced during the writing of this dissertation.

They do not compromise the integrity and originality of this thesis. All writing is original and the result of our work. Information from other authors is properly cited throughout the document.

Sofia Alexandra Santos Soares

June, 2025