

# Interest Rate Forecasting with Dynamic Nelson-Siegel Models

Evaluating Traditional and Distribution-Free Methods for Risk  
Assessment

Ellen Waara Ankarstrand

April 2025

# Abstract

Accurate yield curve forecasts are essential for financial planning and risk management. This thesis evaluated the predictive performance of three variants of the Dynamic Nelson-Siegel (DNS) models: a one-step DNS using the Kalman filter, a two-step DNS with a Vector Autoregressive VAR(1) model, and a simulation-based method using block-bootstrap resampling. To quantify predictive uncertainty, conformal prediction intervals were applied, offering finite-sample coverage guarantees without relying on distributional assumptions.

The models were assessed on their forecast accuracy and ability to replicate key statistical properties of the observed curve dynamics, including autocorrelation and distributional moments. The analysis was conducted using daily interest rate data from 2010 to 2024 across three markets: Stockholm Interbank Offered Rate (Stibor), Euro Interbank Offered Rate (Euribor), and the US federal funds rate.

The results showed that the block-bootstrap method consistently achieved the lowest forecast error and best replicated the statistical properties. The one-step DNS produced the widest prediction intervals, was most sensitive to market volatility, and struggled to accurately replicate the variance and autocorrelation. The two-step DNS performed reasonably well overall, except for the poor estimation of the higher-order moments. Across all models, the conformal prediction intervals delivered empirical coverage well above the nominal level, with the bootstrap method producing the narrowest and most stable intervals.

These findings demonstrate the practical value of combining DNS models with distribution-free uncertainty quantification and offer a comparative view of model strengths and limitations in the context of risk management.

## Keywords

Yield Curve Forecasting, Dynamic Nelson-Siegel Models, Interest Rate Risk, Fixed Income, Conformal Prediction, Block Bootstrap, Kalman Filter, Vector Autoregression (VAR), Risk Management, Forecast Evaluation

## Sammanfattning

Tillförlitliga prognoser för räntor är avgörande för finansiell planering och riskhantering. I denna uppsats utvärderades den prediktiva förmågan hos tre varianter av den dynamiska Nelson-Siegel-modellen (DNS): en enstegsmodell med Kalmanfilter, en tvåstegsmodell med VAR(1), samt en bootstrapbaserad simuleringsmetod. För att konstruera konfidensintervall användes *conformal prediction*, som ger täckningsgarantier utan antaganden om underliggande fördelningar.

Modellerna utvärderades utifrån deras precision samt förmågan att återge centrala statistiska egenskaper hos de observerade räntekurvorna, såsom autokorrelation och moment. Analysen baserades på dagliga referensräntor för Stockholm Interbank Offered Rate (Stibor), Euro Interbank Offered Rate (Euribor) och den amerikanska styrräntan (federal funds rate) under perioden 2010–2024.

Resultaten visade att bootstrapmetoden konsekvent uppnådde högst precision och bäst återgav de statistiska egenskaperna. Enstegsmodellen gav de bredaste prediktionsintervallet, störst fel och hade svårt att återge både varians och autokorrelation. Tvåstegsmodellen presterade överlag väl, men över- respektive underskattade tredje och fjärde ordningens moment. Konfidensintervallet hade en empirisk täckning klart över den nominella nivån på 90%, där bootstrapmetoden gav de smalaste och mest stabila intervallet.

Resultaten visar på det praktiska värdet av att kombinera DNS-modeller med conformal prediction, samt diskuterar modellernas styrkor och svagheter ur ett riskperspektiv.

## Nyckelord

Prediktion av räntekurvor, ränterisk, marknadsrisk, Nelson-Siegel, conformal prediction, block-bootstrap, Kalmanfilter, vektorautoregression (VAR)

## Acknowledgements

I extend my deepest gratitude to my supervisors, Patrik Hellgren and Jonatan Sköld at SEK Svensk Exportkredit, for their direction, encouragement, and constructive feedback. This thesis would not have been possible without their support.

I am equally thankful to my academic supervisor, Dr. Soon Hoe Lim, for his guidance and invaluable suggestions throughout the research process. His perspective and expertise significantly enhanced the quality of this work.

Special thanks go to the Market Risk team at SEK Svensk Exportkredit for their encouragement, insights, and for warmly welcoming me as part of the team.

Finally, I am immensely grateful to my partner, Mattias, for his endless patience, understanding, and moral support.

# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Yield Curves as Economic Barometers . . . . .	7
1.2	The Evolution of Yield Curve Modelling . . . . .	8
1.3	Research Objectives . . . . .	9
<b>2</b>	<b>Financial Preliminaries</b>	<b>11</b>
2.1	Interest Rates . . . . .	11
2.2	Yield Curves . . . . .	12
2.2.1	Properties of Yield Curves . . . . .	12
<b>3</b>	<b>The Nelson-Siegel Model</b>	<b>14</b>
3.1	The Static Nelson-Siegel Model . . . . .	14
3.1.1	Definition . . . . .	14
3.1.2	Interpretation . . . . .	14
3.2	The Dynamic Nelson-Siegel Model . . . . .	15
3.3	State-Space Formulation . . . . .	15
<b>4</b>	<b>Forecasting Methods</b>	<b>18</b>
4.1	Two-Step DNS . . . . .	18
4.1.1	Choosing $\lambda$ . . . . .	18
4.2	One-Step DNS with the Kalman Filter . . . . .	19
4.2.1	Initialization . . . . .	20
4.2.2	Filtering Steps . . . . .	20
4.3	Bootstrap-Based Estimation . . . . .	21
4.3.1	Sampling Method . . . . .	22
4.3.2	Exponential Smoothing . . . . .	22
4.4	Limitations . . . . .	23
4.4.1	Limitations of the 2-step DNS with VAR(1) . . . . .	23

4.4.2	Limitations of the 1-step DNS with Kalman Filter . . . . .	24
4.4.3	Limitations of the Bootstrap DNS . . . . .	24
<b>5</b>	<b>Quantifying Uncertainty</b>	<b>25</b>
5.1	Conformal Prediction Intervals . . . . .	25
5.1.1	The Choice of Split or Full Conformal Prediction . . . . .	27
5.1.2	Non-Exchangeable Conformal Prediction . . . . .	27
5.1.3	Multi-Step Forecasting . . . . .	28
5.2	Empirical Coverage . . . . .	28
5.3	Statistical Properties . . . . .	28
<b>6</b>	<b>Data Analysis</b>	<b>30</b>
6.1	The Dataset . . . . .	30
6.2	Estimation Framework . . . . .	33
<b>7</b>	<b>Results</b>	<b>35</b>
7.1	US Federal Funds . . . . .	35
7.1.1	Conformal Prediction . . . . .	35
7.1.2	Mean Squared Error and Pearson Correlation . . . . .	38
7.1.3	Statistical Properties . . . . .	41
7.2	Euro Interbank Offered Rate (Euribor) . . . . .	47
7.2.1	Conformal Prediction . . . . .	47
7.2.2	Mean Squared Error and Pearson Correlation . . . . .	50
7.2.3	Statistical Properties . . . . .	53
7.3	Stockholm Interbank Offered Rate (Stibor) . . . . .	59
7.3.1	Conformal Prediction . . . . .	59
7.3.2	Mean Squared Error and Pearson Correlation . . . . .	62
7.3.3	Statistical Properties . . . . .	65
7.4	Stability Analysis . . . . .	71
7.4.1	One-Step DNS with Kalman Filter . . . . .	71
7.4.2	Two-Step DNS with VAR(1) . . . . .	71
7.4.3	Bootstrap DNS . . . . .	72
7.4.4	Conformal Prediction . . . . .	72
<b>8</b>	<b>Discussion</b>	<b>73</b>
8.1	Results Summary . . . . .	73
8.1.1	Mean Squared Error and Pearson Correlation . . . . .	73
8.1.2	Conformal Prediction Intervals . . . . .	74
8.1.3	Statistical Properties . . . . .	74

8.1.4	Interpretation of Results . . . . .	75
8.2	Implications for Risk . . . . .	76
8.2.1	Model Comparison . . . . .	76
8.2.2	Takeaways . . . . .	77
8.3	Limitations . . . . .	78
8.4	Future Work . . . . .	79
<b>A</b>	<b>Appendix</b>	<b>80</b>
A.1	State Space Models . . . . .	80
A.2	Kalman Filtering . . . . .	81
<b>References</b>		<b>84</b>

# Chapter 1

## Introduction

### 1.1 Yield Curves as Economic Barometers

Yield curves describe the relationship between interest rates and maturities of government bonds and are critical indicators in macroeconomic analysis and financial decision-making. Their shape reflects the market's collective outlook on growth, inflation, and monetary policy. A steep, upward-sloping curve typically indicates expectations of stronger growth, rising inflation, and higher interest rates, whereas a flat or downward curve implies anticipated rate cuts and economic weakness.

In recent years, the significance of yield curves has been underscored by extraordinary global events, from the prolonged low interest rates after the 2008 financial crisis to the sharp hikes in 2022. Curves flattened, inverted, or otherwise deviated in ways that carried significant implications for economic planning and financial stability. In light of these developments, reliable yield curve modelling remains a central concern for economic analysts, financial institutions, and researchers alike.

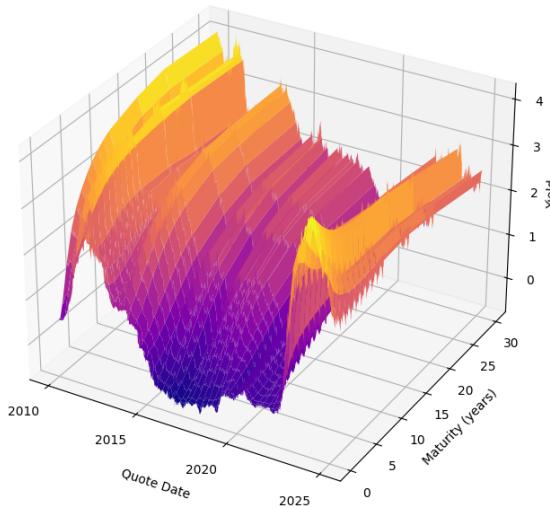


Figure 1.1: Stibor curve over time and maturity.

## 1.2 The Evolution of Yield Curve Modelling

Given the yield curve's significance, researchers have developed various models to describe its shape and dynamics. An important contribution was made by Nelson et al. (1987) who proposed a simple parametric model capable of fitting diverse curve shapes using just three latent factors, suitably called *the Nelson-Siegel model*. These factors are commonly interpreted as the yield curve's level, slope, and curvature. The model became widely adopted in both academia and industry due to its balance of flexibility and simplicity.

While the original Nelson-Siegel model is static — fitting the curve at a single point in time — later research recognised the importance of modelling the curve's evolution over time. F. Diebold and Li (2006) introduced a dynamic version of the model, in which the level, slope, and curvature are treated as time-varying latent factors. In their Dynamic Nelson-Siegel model, often called 2-step DNS, the factors are first estimated from historical yield data and then projected forward using time-series models, enabling forecasts of the entire yield curve.

Around the same time, F. Diebold, Rudebusch, et al. (2006) further developed the model by formulating it in a state-space form and estimating it using the Kalman filter. In this implementation, the latent factors are treated as unobserved state

variables that evolve according to a stochastic process, while the observed yields are modelled as noisy measurements of these states. The Kalman filter applies maximum likelihood to jointly estimate the factors and model parameters, resulting in what is commonly referred to as the 1-step DNS. The 1-step DNS allows for the simultaneous estimation of all parameters, yielding statistically valid inference under standard theory. In contrast, the 2-step procedure does not account for the uncertainty introduced in the first stage of factor estimation (F. Diebold, Rudebusch, et al. 2006).

Since then, further extensions have been proposed, including the five-factor Generalised Dynamic Nelson-Siegel model (Christensen, F. X. Diebold, et al. 2009) and the Arbitrage-Free Nelson-Siegel model (Christensen, F. Diebold, et al. 2011), which imposes no-arbitrage conditions to align the model with financial theory.

### 1.3 Research Objectives

This thesis addresses three main objectives related to yield curve forecasting using the Dynamic Nelson-Siegel (DNS) framework.

The first objective is to compare the predictive performance of three DNS-based forecasting approaches. Specifically, the study evaluates: (1) a one-step DNS model estimated via the Kalman filter, (2) a two-step DNS model using a vector autoregressive (VAR) model, and (3) a simulation-based method that combines DNS with a moving-block bootstrap. The models are assessed based on their forecasting accuracy, stability, and ability to preserve the underlying statistical properties of yield curves.

The second objective is to improve uncertainty quantification by applying conformal prediction to the DNS forecasts. Conformal prediction is a distribution-free method for generating prediction intervals with theoretical coverage guarantees. Given its limited application in finance, this study aims to contribute new insights into the potential of conformal prediction in financial forecasting and risk assessment.

The third objective is to apply these forecasting techniques to interest rate data from three markets: Sweden (Stibor), the euro area (Euribor), and the United States (US federal funds rate). While much of the existing literature focuses on U.S. Treasury yields, smaller or region-specific markets are comparatively underexplored. This study provides a broader model performance evaluation by including multiple currencies.

This analysis is motivated by the needs of Svensk Exportkredit (SEK), a state-owned company that provides long-term financing to Swedish exporters and their international customers. As SEK is exposed to interest rate risks across all three currency areas, this study intends to support decision-making related to risk management.

# Chapter 2

## Financial Preliminaries

### 2.1 Interest Rates

Following Bohner (2011), the following concepts are relevant when working with interest rates.

**Definition 2.1.1** (Zero-coupon bond). *A zero-coupon bond with maturity  $T > 0$  is a contract that guarantees the holder a single payment of one unit of currency at the future time  $T$ . The price of such a bond at any time  $t \in [0, T]$  is denoted by  $P(t, T)$ . At time  $t$ , the time to maturity is  $T - t$ , or more generally, when accounting for day-count conventions,  $\tau(t, T)$ .*

**Definition 2.1.2** (Forward rate). *The forward rate for the period  $[T, S]$  is the continuously-compounded interest rate agreed upon at time  $t$  for lending or borrowing over the future period from  $T$  to  $S$ . It is given by*

$$R(t; T, S) = -\frac{\ln P(t, S) - \ln P(t, T)}{\tau(T, S)},$$

where  $P(t, T)$  and  $P(t, S)$  denote the prices at time  $t$  of zero-coupon bonds maturing at  $T$  and  $S$ , respectively, and  $\tau(T, S)$  is the time (in years) between  $T$  and  $S$ , adjusted for day-count conventions.

**Definition 2.1.3** (Continuously-compounded spot interest rate). *The continuously-*

compounded spot interest rate with maturity  $T$  prevailing at time  $t$  is defined as

$$R(t, T) = -\frac{\ln P(t, T)}{\tau(t, T)},$$

where  $P(t, T)$  is the price at time  $t$  of a zero-coupon bond maturing at  $T$ , and  $\tau(t, T)$  is the time to maturity. It is also called the **yield to maturity**.

**Definition 2.1.4** (Simply-compounded spot interest rate). The simply-compounded spot interest rate with maturity  $T$  prevailing at time  $t$  is defined as

$$L(t, T) = \frac{1 - P(t, T)}{\tau(t, T)P(t, T)},$$

where  $P(t, T)$  is the price at time  $t$  of a zero-coupon bond maturing at  $T$ , and  $\tau(t, T)$  is the time to maturity.

## 2.2 Yield Curves

Yield curves describe the relationship between time to maturity and the interest rates associated with zero-coupon bonds at a given time. Formally, it is the mapping  $T \mapsto R(t, T)$ , where  $R(t, T)$  is the continuously-compounded spot rate prevailing at time  $t$  for maturity  $T$ .

In practice, prices of zero-coupon bonds are only available for a limited set of maturities. As a result, yield curves must be estimated from observable market instruments such as government bonds, interest rate swaps, and futures. The resulting curve reflects the market's pricing of the time value of money and embeds expectations about future interest rates, risk, and liquidity conditions.

Yield curves evolve over time and across maturities, forming a three-dimensional surface. The cross-sectional view captures the term structure at a given time, showing how yields vary with maturity  $T$ . The temporal view shows how the yield curve evolves over calendar time  $t$ .

### 2.2.1 Properties of Yield Curves

Empirical yield curves exhibit several essential properties. First, average yields tend to increase with maturity as a compensation for duration risk. Second, yield volatility

generally decreases with maturity. Short-term interest rates are more sensitive to monetary policy decisions and macroeconomic news, which makes them more volatile. Finally, yields, particularly those at longer maturities, are strongly autocorrelated, meaning that current yield levels resemble yields in the recent past.

# Chapter 3

## The Nelson-Siegel Model

### 3.1 The Static Nelson-Siegel Model

#### 3.1.1 Definition

The Nelson-Siegel model, first described by Nelson et al. (1987), is a three-factor parametric model. It approximates the forward rate curve (which describes the instantaneous interest rate) at a time  $t$  using

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda e^{-\lambda\tau},$$

where  $\beta_i$  are latent parameters to be estimated,  $\lambda$  rate of decay,  $\tau$  time to maturity.

By integrating the forward rate curve, the **static Nelson-Siegel Yield** curve is obtained:

$$\begin{aligned} y(\tau) &= \frac{1}{\tau} \int_0^\tau f(u) du \\ &= \beta_1 + \beta_2 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \end{aligned}$$

where  $y(\tau)$  is the zero-coupon yield curve with  $\tau$  months to maturity.

#### 3.1.2 Interpretation

The parameter  $\lambda$  controls the rate of exponential decay, where a small  $\lambda$  produces a slow decay suited for long-term maturities, and vice versa.

According to F. Diebold and Li (2006) the parameters  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  can be interpreted as three latent, dynamic factors:

- $\beta_1$  (**level**): Has constant loading, meaning it does not decay. It represents the long-term factor that determines the level of the yield curve.
- $\beta_2$  (**slope**): Its loading starts at one and quickly decays to 0, capturing short-term effects and the yield curve's slope.
- $\beta_3$  (**curvature**): Its loading starts at 0, increases, and then decays back to 0. It is the medium-term factor related to the curve's curvature.

## 3.2 The Dynamic Nelson-Siegel Model

Building on the work of Nelson et al. 1987, the model was extended to a dynamic framework by F. Diebold and Li 2006. Here, the  $\beta_i$ :s are no longer fixed parameters but time-dependent variables, allowing the model to adapt as the yield curve changes over time. This gives us the *Dynamic Nelson-Siegel (DNS)* model:

$$\begin{aligned} y_t(\tau) &= \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \\ &= L_t + S_t\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \end{aligned}$$

Here,  $\beta_{1t}, \beta_{2t}, \beta_{3t}$  or  $L_t, S_t, C_t$  are time-series of latent parameters with the same interpretation as level, slope and curve as in the static Nelson-Siegel model. On the contrary, the parameter  $\lambda$  is considered an inherent property of the model and does not vary with  $t$ .

## 3.3 State-Space Formulation

The dynamic Nelson-Siegel model can be interpreted as a state-space model:

$$\begin{aligned} (f_t - \mu) &= A(f_{t-1} - \mu) + \eta_t, \\ y_t &= \Lambda f_t + \varepsilon_t. \end{aligned}$$

The errors  $\eta_t$  are static shocks and  $\varepsilon_t$  are measurement disturbances that follow

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \sim WN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \right).$$

We require  $\eta_t$  and  $\varepsilon_t$  to be orthogonal (uncorrelated) and the initial state to be:  $\mathbb{E}[f_0\eta'_t] = 0$ ,  $\mathbb{E}[f_0\varepsilon'_t] = 0$  (F. Diebold, Rudebusch, et al. (2006)) .

### The transition equation

The transitions equation describes how the latent factors evolve over time:

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t,$$

where

$$f_t = \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_L \\ \mu_S \\ \mu_C \end{pmatrix}.$$

The model can take two forms: *independent-factor DNS* and *correlated-factor DNS*. The independent-factor DNS assumes that each factor follows a first-order autoregressive process,  $AR(1)$ , leading to a diagonal state transition matrix  $A$  and uncorrelated shocks  $\eta_t$  with a diagonal covariance matrix:

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}, \quad Q = \begin{pmatrix} q_{11}^2 & 0 & 0 \\ 0 & q_{22}^2 & 0 \\ 0 & 0 & q_{33}^2 \end{pmatrix}.$$

In contrast, the correlated-factor DNS assumes that each factor follows a first-order vector-autoregressive process,  $VAR(1)$ , resulting in a full transition matrix and correlated shocks, meaning the covariance matrix is non-diagonal:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad Q = qq^\top, \quad q = \begin{pmatrix} q_{11} & 0 & 0 \\ q_{21} & q_{22} & 0 \\ q_{31} & q_{32} & q_{33} \end{pmatrix}.$$

Independent-factor DNS is usually assumed, meaning that the shocks to one factor do not affect the others. Independent factors correspond to uncorrelated deviations in yields across maturities, which is a common assumption in no-arbitrage term structure models where independent measurement errors are added to observed yields. It also improves computational tractability when working with a large number of maturities.

## The measurement equation

The measurement equation relates the observed yields as a function of the three latent factors:

$$y_t = \Lambda f_t + \varepsilon_t.$$

For a set of  $N$  yields and  $t = 1, \dots, T$  we have

$$y = \begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix}, \quad f_t = \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix}.$$

The factor load matrix  $\Lambda \in \mathbb{R}^{N \times 3}$  is given by

$$\Lambda = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix}.$$

# Chapter 4

## Forecasting Methods

### 4.1 Two-Step DNS

The two-step method is a simple and numerically stable way to fit the Nelson-Siegel model. The standard practice, proposed by Nelson et al. (1987), is as follows:

The first step involves fitting the static Nelson-Siegel model to each time point  $t = 1, \dots, T$ . First,  $\lambda$  is fixed to a pre-specified value. The factor loadings  $f$  are computed for each maturity using this fixed value. Then, for each time point  $t$ , ordinary least squares (OLS) is used to estimate the three Nelson-Siegel factors, resulting in a time series  $\{L_t, S_t, C_t\}_{t=1}^T$ .

In the second step, the dynamics of the estimated factors are modelled using first-order vector autoregression,  $VAR(1)$ , model. Formally, letting  $f_t = \{L_t, S_t, C_t\}_{t=1}^T$ , the model is

$$f_t = A + B f_{t-1} + \varepsilon_t, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

where  $A \in \mathbb{R}^3$  is the offset vector and  $B_i \in \mathbb{R}^{3 \times 3}$  is the VAR coefficient matrix.

#### 4.1.1 Choosing $\lambda$

There are multiple ways to determine an appropriate value for  $\lambda$ . A common choice, suggested by F. Diebold and Li (2006), is setting it to  $\lambda = 0.0609$ . The motivation for this choice lies in the model's structure. The parameter  $\lambda$  determines the maturity at which the loading on the curvature factor,  $C_t$ , peaks. Since two- to three-year

maturities are often considered representative of the medium-term segment of the yield curve, Diebold and Li suggest setting their midpoint, 30 months, as a reference. The value  $\lambda = 0.0609$  is chosen to maximise the loading on the curvature factor at 30 months.

## 4.2 One-Step DNS with the Kalman Filter

*Details on the Kalman filter are covered in the appendix.*

As previously shown in 4.2, the state space representation of the Dynamic Nelson-Siegel model is:

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t, \\ y_t = \Lambda f_t + \varepsilon_t,$$

where, for a set of  $N$  maturities,

- $f_t = (L_t, S_t, C_t)^\top$  is the latent state vector
- $\mu \in \mathbb{R}^3$  is the vector of factor means
- $A \in \mathbb{R}^{3 \times 3}$  is the transition matrix
- $\Lambda \in \mathbb{R}^{N \times 3}$  is the factor loading matrix (depends on the decay parameter  $\lambda$ )
- $\eta_t \sim \mathcal{N}(0, Q)$ ,  $Q \in \mathbb{R}^{3 \times 3}$  is the transition noise
- $\varepsilon_t \sim \mathcal{N}(0, R)$ ,  $R \in \mathbb{R}^{N \times 3}$  is the measurement noise

The error terms are assumed to follow:

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \sim WN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \right), \\ \mathbb{E}[f_0 \eta'_t] = 0, \quad \mathbb{E}[f_0 \varepsilon'_t] = 0.$$

Before applying the Kalman filter, we rewrite the model in standard state-space form by expanding the transition equation:

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t, \\ \Rightarrow f_t = Af_{t-1} + (I - A)\mu + \eta_t$$

Set  $f_t = x_t$  to follow standard notation:

$$\begin{aligned}x_t &= (I - A)\mu + Ax_{t-1} + \eta_t \\y_t &= \lambda x_t + \varepsilon_t.\end{aligned}$$

The product  $(I - A)\mu$  is called the *offset* of A.

### 4.2.1 Initialization

Following F. Diebold, Rudebusch, et al. (2006), the filter is initialised by setting the initial state mean to  $X_0 = \mu$  and solving the initial covariance from the discrete Lyapunov equation:

$$P_0 = V \quad \text{where } V = AVA^\top + Q.$$

The transition matrix  $A$  is estimated through least squares regression using a sample of historical states. The residuals from this regression are used to estimate the process noise covariance  $Q$ . The decay parameter is initialized to  $\lambda = 0.0609$ , as recommended by F. Diebold, Rudebusch, et al. (2006)

### 4.2.2 Filtering Steps

For each time step  $t$ , perform the prediction and update steps:

#### Prediction step:

$$\begin{aligned}x_{t|t-1} &= (I - A)\mu + Ax_{t-1} \\P_{t|t-1} &= AP_{t-1}A^\top + Q\end{aligned}$$

Where 4.2.2 is the predicted state and 4.2.2 is the predicted covariance.

#### Update step

$$\begin{aligned}u_t &= Y_t - \Lambda X_{t|t-1} \\F_t &= cov(u_t) = \Lambda P_{t|t-1} \Lambda^\top + R \\K_t &= P_{t|t-1} \Lambda^\top F_t^{-1} \\X_t &= X_{t|t-1} + K_t u_t \\P_t &= (I - K_t \Lambda) P_{t|t-1}\end{aligned}$$

Here,  $u_t$  is the innovation/residual,  $F_t$  is its covariance,  $K_t$  is the Kalman gain, and  $x_t, P_t$  are the updated state and covariance.

**Log-Likelihood** Given the full set of parameters, the log-likelihood of the model is computed using the prediction error decomposition:

$$\log L = -\frac{TN}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T (u_t^\top F_t^{-1} u_t + \log |F_t|)$$

where  $T$  is the number of time steps.

**Likelihood Maximisation and Forecasting** Using the likelihood, maximum likelihood estimates of the parameters,  $\delta_{ML}$ , are computed numerically.

### 4.3 Bootstrap-Based Estimation

This bootstrap-based estimation simulates a future time series of model parameters by resampling past parameter changes from a historical calibration window. Instead of imposing a parametric model for future dynamics, it generates a future sequence of parameters by drawing from past changes.

At a time  $t$ , let  $N$  be the window length, and define the calibration window as  $(t - N, t)$ , containing  $N + 1$  past observations. Define

- $p_i$  for  $i \in (t - N, t)$ : the estimated model parameters.
- $\Delta p_i = p_i - p_{i-1}$  for  $i \in (t - N + 1, t)$ : the changes in model parameters.

The goal is to construct a time series of future parameters for  $d$  starting at time  $t$ , which is done by repeatedly resampling segments of past changes with stochastic length

$$\{\Delta p_{t_{start}^i}, \dots, \Delta p_{t_{end}^i}\}$$

until their combined length exceeds  $d$ . That is, segments are sampled until

$$\{\Delta p_{t_{start}^1}, \dots, \Delta p_{t_{end}^1}\}, \dots, \{\Delta p_{t_{start}^n}, \dots, \Delta p_{t_{end}^n}\}$$

contains more than  $d$  observations. The future parameter sequence is then given by

$$p(t + \Delta t) = p(t) + \sum_{j \leq d} \Delta p_{t_j}.$$

### 4.3.1 Sampling Method

1. Sample a window length  $l \sim U(L_{min}, L_{max})$ .
2. Select an endpoint  $t_{end}$ 
  - Randomly sample  $t_{end}$  from the interval  $(t - N + 1 + L_{max}, t)$ .
  - Weights that increase the probability of selecting more recent values are applied.
3. Compute a starting point:  $t_{start} = t_{end} - l$
4. Form a window by extracting the sequence  $\{\Delta p_{t_{start}^i}, \dots, \Delta p_{t_{end}^i}\}$

**Sampling  $t_{end}$**  The valid range for  $t_{end}$  is

$$t_{end} \in (t_{min}, t_{max}) = (t - N + 1 + L_{max}, t).$$

A higher probability is assigned to more recent candidates for  $t_{end}$  to reflect the assumption that upcoming changes are more likely to resemble those closer in time. This is done by assigning exponentially increasing weights to each endpoint candidate:

$$w_i = \exp\left(\frac{i}{N}\right), \quad i = 0, 1, \dots, N - 1,$$

where  $N$  is the total number of endpoints. These weights are then normalised to form a probability distribution:

$$p_i = \frac{w_i}{\sum_{j=0}^{N-1} w_j}, \quad \text{such that} \quad \sum_{i=0}^{N-1} p_i = 1.$$

### 4.3.2 Exponential Smoothing

Exponential weighted moving average smoothing was applied to the time series of past parameters before computing their differences,  $\Delta p$ . Weighing reduces short-term fluctuations in the past parameters, so the resulting differences reflect broader trends rather than noise.

The exponentially smoothed parameters,  $\tilde{p}$ , were calculated recursively:

$$\begin{aligned} \tilde{p}_0 &= p_0 \\ \tilde{p}_t &= (1 - \alpha)\tilde{p}_{t-1} + \alpha p_t, \end{aligned}$$

where  $\alpha$  is the smoothing factor.

Figure 4.1 shows an example of parameter smoothing.

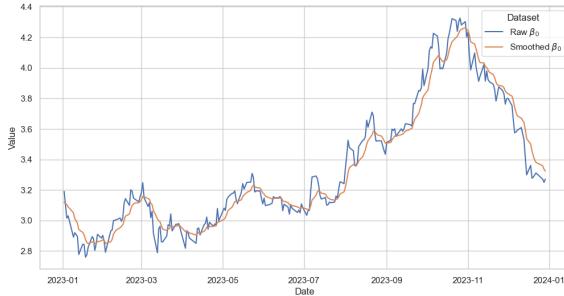


Figure 4.1: Raw vs. smoothed parameter  $\beta_0$  for 2023, using  $\alpha = 0.2$ .

## 4.4 Limitations

All predictive models are constrained by the information available at the time of estimation. Regarding interest rate forecasting, exogenous information—such as monetary policy, geopolitical developments, or changes in market sentiment—remains outside the model’s direct scope. Even when markets anticipate such changes, the Nelson-Siegel model may still struggle to capture them due to its rigid functional form. This can lead to large residuals and weak forecasting performance, especially for shorter maturities that are highly sensitive to external factors, and longer forecast horizons, where considerable developments can occur between the forecast start and target date.

### 4.4.1 Limitations of the 2-step DNS with VAR(1)

The primary drawback of the 2-step DNS model is that it treats the factor estimates from the first step as observed data when fitting the time series in the second step. As a result, estimation errors from the fitting stage do not propagate into the forecasting stage, potentially leading to biased forecasts and underestimated uncertainty. Essentially, one assumes the estimated factors are the "true" factors, ignoring estimation error.

Another issue is that fixing the decay factor  $\lambda$  (often done to linearise the first step) can introduce misspecification if the chosen  $\lambda$  is not optimal for the entire forecasting period.

#### **4.4.2 Limitations of the 1-step DNS with Kalman Filter**

The one-step DNS model addresses the issue with non-propagating errors by estimating the entire system simultaneously within a state-space framework using the Kalman filter. Nonetheless, the state-space formulation increases the model complexity due to the substantial number of parameters. In addition, the Kalman filter is known to be sensitive to initial values and model specification. From a practical perspective, the implementation requires more time, technical expertise, and computational power, which may not always be justified.

#### **4.4.3 Limitations of the Bootstrap DNS**

While the bootstrap method avoids explicit parametric modelling, which is flexible and avoids distributional assumptions, it also has several limitations. First, it does not distinguish between signal and noise in the factor innovations (parameter changes), which can result in less informative forecasts, especially during times of change. Second, the method's performance depends critically on the choice of calibration window. If the historical sample does not reflect the movements and volatility of the forecasting period, the model will be unable to produce plausible results.

# Chapter 5

## Quantifying Uncertainty

### 5.1 Conformal Prediction Intervals

Conformal prediction is a method for constructing prediction intervals around a pre-trained model. It converts any heuristic uncertainty measure from an arbitrary model into a rigorous, statistically valid prediction set. It makes no assumptions about whether the underlying task is discrete or continuous, or classification versus regression (Angelopoulos et al. (2021)).

There are two main types of conformal prediction: split and full (Barber et al. 2023).

**Definition 5.1.1** (Split Conformal Prediction). *Given a model  $\hat{\mu} : \mathcal{X} \rightarrow \mathbb{R}$  fitted on a training set, and given  $n$  additional data points  $(X_1, Y_1), \dots, (X_n, Y_n)$  (the holdout set), define residuals*

$$R_i = |Y_i - \hat{\mu}(X_i)|, \quad i = 1, \dots, n.$$

*The  $(1 - \alpha)$  prediction interval for the new feature vector  $X_{n+1}$  is*

$$\hat{C}_n(X_{n+1}) = \hat{\mu}(X_{n+1}) \pm (\text{the } \lceil (1 - \alpha)(n + 1) \rceil\text{-th smallest of } R_1, \dots, R_n).$$

*Or equivalently*

$$\hat{C}_n(X_{n+1}) = \hat{\mu}(X_{n+1}) \pm Q_{1-\alpha} \left( \sum_{i=1}^n \frac{1}{n+1} \cdot \delta_{R_i} + \frac{1}{n+1} \cdot \delta_{+\infty} \right),$$

where  $Q_\tau(\cdot)$  denotes the  $\tau$ -quantile of its argument, and  $\delta_a$  denotes the Dirac delta function at  $a$ . This method guarantees distribution-free predictive coverage at the target level  $1 - \alpha$ .

**Definition 5.1.2** (Full Conformal Prediction). *The full conformal prediction is applicable to any regression algorithm*

$$\mathcal{A} : \bigcup_{n \geq 0} (\mathcal{X} \times \mathbb{R})^n \rightarrow \{\text{measurable functions } \hat{\mu} : \mathcal{X} \rightarrow \mathbb{R}\},$$

mapping a dataset of input-output pairs  $(X_i, Y_i)$  to a fitted regression function  $\hat{\mu}$ . The output  $\hat{\mu}$  must be invariant under any input permutation, meaning the algorithm  $\mathcal{A}$  treats the data symmetrically.

Given training data  $(X_1, Y_1), \dots, (X_n, Y_n)$  and a test input  $X_{n+1}$ , for each  $y \in \mathbb{R}$  append  $(X_{n+1}, y)$  to the dataset, retrain the model to obtain a new function  $\hat{\mu}^y$ , and then compute the residuals

$$\begin{aligned} R_i^{(y)} &= |Y_i - \hat{\mu}^y(X_i)|, \quad i = 1, \dots, n, \\ R_{n+1}^{(y)} &= |y - \hat{\mu}^y(X_{n+1})|. \end{aligned}$$

The  $p$ -value for  $y$  is given by

$$\pi(y) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{1} \left\{ R_i^{(y)} \geq R_{n+1}^{(y)} \right\},$$

The prediction set for a feature vector  $X_{n+1}$  is

$$\hat{C}_n(X_{n+1}) = \left\{ y \in \mathbb{R} : R_{n+1}^y \leq Q_{1-\alpha} \left( \sum_{i=1}^{n+1} \frac{1}{n+1} \cdot \delta_{R_i^y} \right) \right\}.$$

This procedure guarantees finite-sample coverage at level  $1 - \alpha$ , assuming the data are exchangeable.

**Theorem 1** (Full conformal prediction, Vovk et al. (2005)). *If the data points  $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$  are i.i.d. (or more generally, exchangeable), and the algorithm  $\mathcal{A}$  treats the input data symmetrically as in (6), then the full conformal prediction set defined in (8) satisfies*

$$\mathbb{P} \left( Y_{n+1} \in \hat{C}_n(X_{n+1}) \right) \geq 1 - \alpha.$$

The same result holds for split conformal prediction.

### 5.1.1 The Choice of Split or Full Conformal Prediction

Split conformal prediction was used to construct prediction intervals. While full conformal prediction often yields narrower intervals by avoiding data splitting, it comes with a steep computational cost, as it requires retraining the model for each  $y \in \mathbb{R}$ . Given the relatively large size of the available dataset, the loss of accuracy from sample splitting was considered acceptable.

### 5.1.2 Non-Exchangeable Conformal Prediction

The validity of conformal prediction is built on two main assumptions: that the data points are exchangeable, and that the model fitting algorithm treats them symmetrically. Unfortunately, both assumptions are often violated in the context of time series. As the data distribution shifts over time, the exchangeability assumption does not hold. Moreover, a non-symmetric algorithm is often preferred in such a setting, giving more weight to recent observations.

Barber et al. (2023) modifies the conformal prediction algorithm to make it more suitable for time series data. They incorporate weighted quantiles to enhance robustness against distribution shifts and propose a randomisation technique that allows for non-symmetric treatment of the data points.

**Weighted Quantiles** Assign weights  $w_1, \dots, w_n \in [0, 1]$  to each data point, with a higher weight assigned to data points  $Z_i$  believed to be closer in distribution to the test point  $Z_{n+1}$ . In a time series setting, a concern could be distribution drift, in which case the weights can be chosen such that the prediction interval relies mainly on recent data points. The weights are assumed to be fixed.

To simplify notation, the weights are normalised. Given  $w_i \in [0, 1], i = 1, \dots, n$ , the normalised weights are defined as

$$\begin{aligned}\tilde{w}_i &= \frac{w_i}{w_1 + \dots + w_n + 1}, \quad i = 1, \dots, n, \\ \tilde{w}_{n+1} &= \frac{1}{w_1 + \dots + w_n + 1}.\end{aligned}$$

**Definition 5.1.3** (Non-exchangeable split conformal with a symmetric or non-symmetric algorithm). *The prediction interval is given by*

$$\widehat{C}_n(X_{n+1}) = \hat{\mu}(X_{n+1}) \pm Q_{1-\alpha} \left( \sum_{i=1}^n \tilde{w}_i \cdot \delta_{R_i} + \tilde{w}_{n+1} \cdot \delta_{+\infty} \right),$$

where  $R_i = |Y_i - \hat{\mu}(X_i)|$  for the pre-trained model  $\hat{\mu}$ , as before.

The parameter  $\rho \in (0, 1]$  is the geometric decay factor used to compute the weights. When  $\rho = 1$ , the weights are uniform, corresponding to the exchangeable case.

### 5.1.3 Multi-Step Forecasting

The split conformal method described above targets single-step forecasting. A rolling-window approach with overlapping blocks was used to extend it to multi-step forecasting.

To construct a  $(1 - \alpha)$  prediction interval for time  $t$ , 11 was used with residuals  $R_i$ ,  $i \in (t - w, t)$ . Then, by rolling  $t$  forward, the interval was allowed to adapt dynamically to the most recent observations.

## 5.2 Empirical Coverage

The empirical coverage of a prediction interval  $\widehat{C}_n(X_{n+1})$  is the proportion of samples for which the estimated value  $\hat{\mu}^y(X_n)$  falls inside  $\widehat{C}_n(X_{n+1})$ . It is calculated as

$$\hat{\tau}_n = \frac{c}{N},$$

where  $c$  is the number of times the estimated value  $\hat{\mu}^y(X_n)$  falls inside the interval  $\widehat{C}_n(X_{n+1})$ , and  $N$  is the total number of samples.

This study calculates the empirical coverage per maturity, with the total number of samples being the whole time series.

## 5.3 Statistical Properties

The following statistical properties were used to evaluate the yield curve models' forecasting accuracy.

**Mean Squared Error (MSE)** Mean squared error measures the average squared difference between observed yields  $Y_i$  and predicted yields  $\hat{Y}_i$ . It is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

where  $n$  denotes the number of observations.

**Pearson Correlation Coefficient (PCC)** The Pearson correlation coefficient measures the strength and direction of the linear relationship between two datasets. It is computed as the covariance of two variables normalised by their standard deviations, resulting in a value between  $-1$  and  $1$ .  $1$  indicates perfect linear correlation between X and Y, and  $0$  signifies no relationship.

**Autocorrelation** Autocorrelation measures the correlation of a time series with its own lagged values. In this study, autocorrelation was computed for both the observed and forecasted yields to examine whether the forecasted time series preserves the persistence patterns of the actual data. A close match suggests that the model captures underlying temporal dynamics, rather than simply fitting pointwise values.

**Moments** The moments of function are measures related to the shape of the curve and include mean, variance, skewness, and kurtosis.

# Chapter 6

## Data Analysis

### 6.1 The Dataset

The dataset consists of continuously compounded spot interest rate curves derived from three reference rates: the Federal Funds rate, the Euro Interbank Offered Rate (Euribor) 3M, and the Stockholm Interbank Offered Rate (Stibor) 3M. The curves were derived as discounting rate curves from market prices of associated instruments, such as futures and interest rate swaps, on these reference rates. The day count convention applied for these rate curves is Act/365. The derived curves were provided by SEK Svensk Exportkredit.

The data are observed daily from 2010-06-09 to 2024-12-09. Maturities include 3 months, 6 months, 1–10 years (annually), 15 years, 20 years, and 30 years.

Figure 6.1 presents three-dimensional surface plots of yield curves for each reference rate over time. Summary statistics—including the mean, standard deviation, minimum, maximum, and autocorrelations at lags 1 and 12—are provided in tables 6.1, 6.2, and 6.3.

Table 6.1: Descriptive statistics of the US federal funds rate.

Maturity (Y)	Mean (%)	St. D.	Min (%)	Max (%)	$\rho(1)$	$\rho(12)$
0.25	1.304	1.746	0.045	5.475	0.99996	0.99829
0.50	1.327	1.738	0.036	5.510	0.99991	0.99788
1.00	1.347	1.672	0.002	5.495	0.99979	0.99675
2.00	1.401	1.466	-0.035	5.034	0.99956	0.99425
3.00	1.497	1.297	-0.030	4.780	0.99934	0.99173
4.00	1.623	1.170	0.006	4.657	0.99884	0.98884
5.00	1.752	1.076	0.041	4.596	0.99887	0.98677
6.00	1.859	1.009	0.095	4.560	0.99869	0.98499
7.00	1.960	0.958	0.154	4.535	0.99856	0.98344
8.00	2.049	0.922	0.210	4.518	0.99845	0.98230
9.00	2.111	0.895	0.261	4.510	0.99834	0.98144
10.00	2.224	0.894	0.305	4.506	0.99834	0.98150
15.00	2.444	0.859	0.433	4.530	0.99821	0.98030
20.00	2.543	0.846	0.447	4.471	0.99819	0.98028
30.00	2.552	0.815	0.436	4.538	0.99812	0.97993

Table 6.2: Descriptive statistics of the Euribor rate.

Maturity (Y)	Mean (%)	St. D.	Min (%)	Max (%)	$\rho(1)$	$\rho(12)$
0.25	0.553	1.317	-0.613	4.032	0.99995	0.99864
0.50	0.565	1.316	-0.577	4.028	0.99992	0.99828
1.00	0.591	1.293	-0.583	3.989	0.99982	0.99706
2.00	0.635	1.221	-0.642	3.815	0.99963	0.99471
3.00	0.714	1.179	-0.655	3.592	0.99955	0.99367
4.00	0.805	1.154	-0.645	3.426	0.99949	0.99307
5.00	0.909	1.143	-0.621	3.374	0.99943	0.99269
6.00	1.015	1.136	-0.584	3.359	0.99941	0.99249
7.00	1.120	1.131	-0.545	3.355	0.99939	0.99229
8.00	1.221	1.126	-0.499	3.443	0.99937	0.99208
9.00	1.317	1.121	-0.450	3.547	0.99935	0.99188
10.00	1.406	1.119	-0.401	3.636	0.99932	0.99167
15.00	1.731	1.111	-0.233	4.009	0.99929	0.99100
20.00	1.831	1.071	-0.153	4.108	0.99921	0.99032
30.00	1.777	0.983	-0.262	3.899	0.99904	0.98902

Table 6.3: Descriptive statistics of the Stibor rate.

Maturity (Y)	Mean (%)	St. D.	Min (%)	Max (%)	$\rho(1)$	$\rho(12)$
0.25	0.929	1.415	-0.682	4.198	0.99992	0.99812
0.50	0.909	1.414	-0.648	4.240	0.99994	0.99824
1.00	0.974	1.382	-0.588	4.234	0.99985	0.99712
2.00	1.059	1.299	-0.493	4.120	0.99968	0.99470
3.00	1.165	1.225	-0.368	3.890	0.99957	0.99293
4.00	1.274	1.166	-0.243	3.715	0.99947	0.99159
5.00	1.387	1.119	-0.175	3.721	0.99938	0.99042
6.00	1.499	1.082	-0.124	3.790	0.99932	0.98964
7.00	1.601	1.050	-0.074	3.831	0.99926	0.98904
8.00	1.689	1.022	-0.018	3.857	0.99921	0.98848
9.00	1.767	0.996	0.037	3.878	0.99916	0.98789
10.00	1.833	0.973	0.093	3.897	0.99911	0.98732
15.00	2.071	0.901	0.231	4.109	0.99893	0.98502
20.00	2.161	0.859	0.298	4.136	0.99878	0.98349
30.00	2.135	0.832	0.242	4.086	0.99870	0.98296

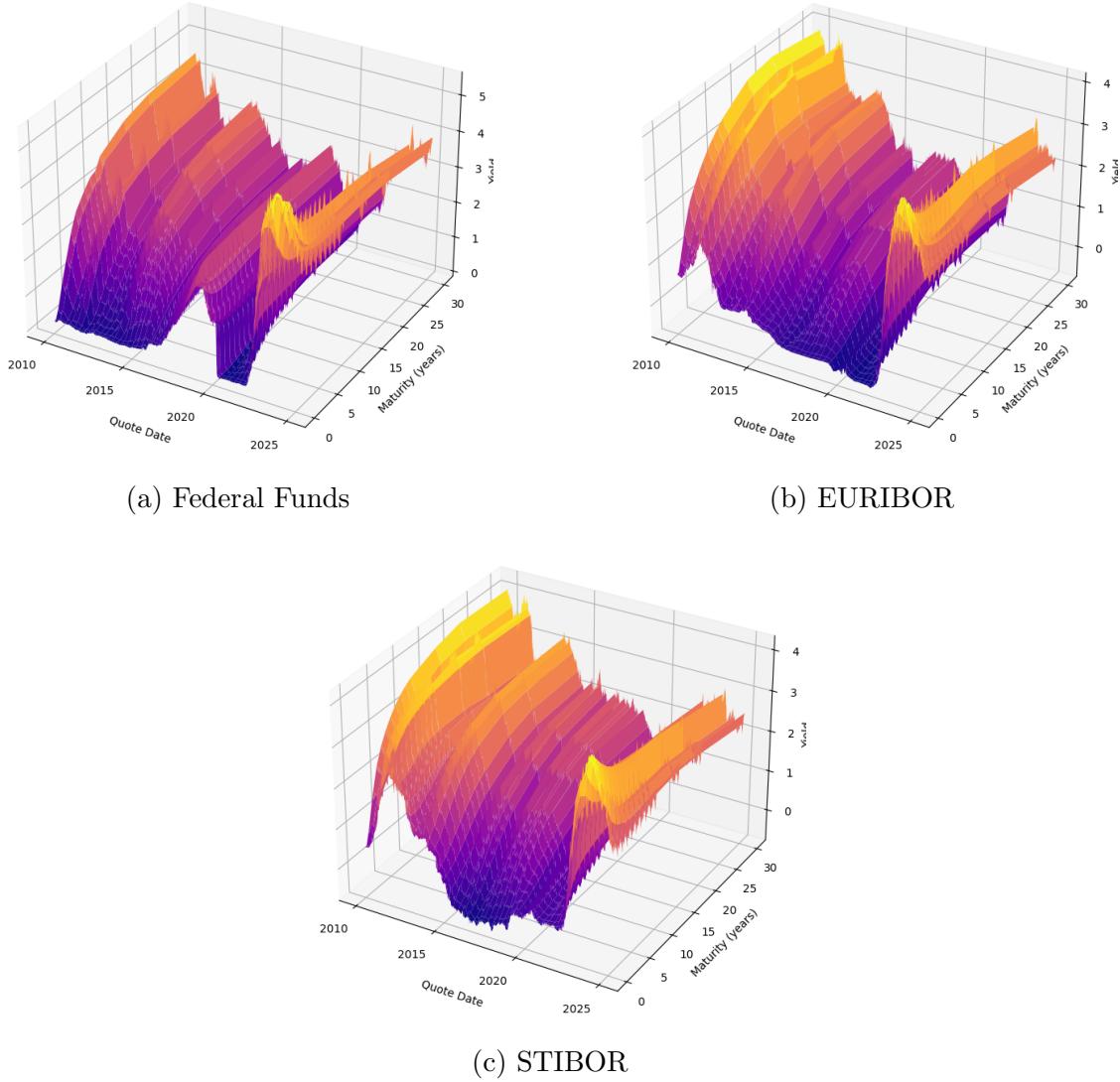


Figure 6.1: Surface plots of the yield curves from 2010–2025.

## 6.2 Estimation Framework

This study evaluates three versions of the Dynamic Nelson–Siegel (DNS) model: a one-step DNS with Kalman filtering, a two-step DNS with a Vector Autoregression

of order one (VAR(1)), and a bootstrapping-based approach. Conformal prediction intervals are constructed alongside standard evaluation metrics such as mean squared error (MSE) and Pearson correlation coefficient (PCC) to assess forecast uncertainty. In addition, the temporal and distributional properties of observed and predicted yields are compared using autocorrelation analysis and moment matching. A rolling window procedure generated 14-day-ahead forecasts, using an estimation window of 360 bank days.

The methods are outlined in chapters 4 and 5.

# Chapter 7

## Results

### 7.1 US Federal Funds

#### 7.1.1 Conformal Prediction

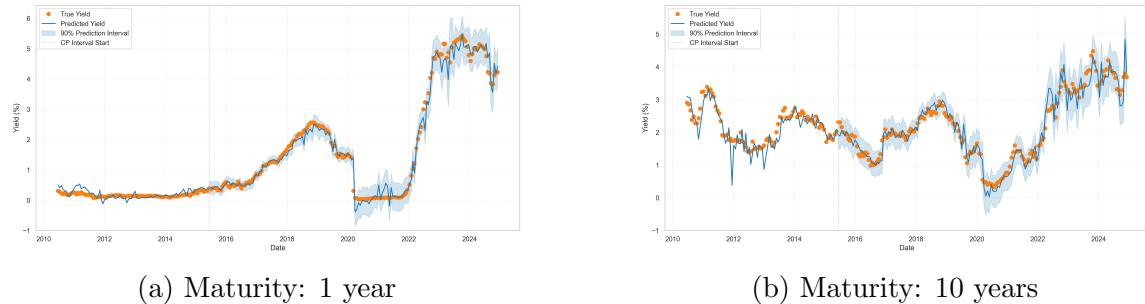


Figure 7.1: One-step DNS Kalman filter forecast with conformal prediction intervals for federal funds. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .

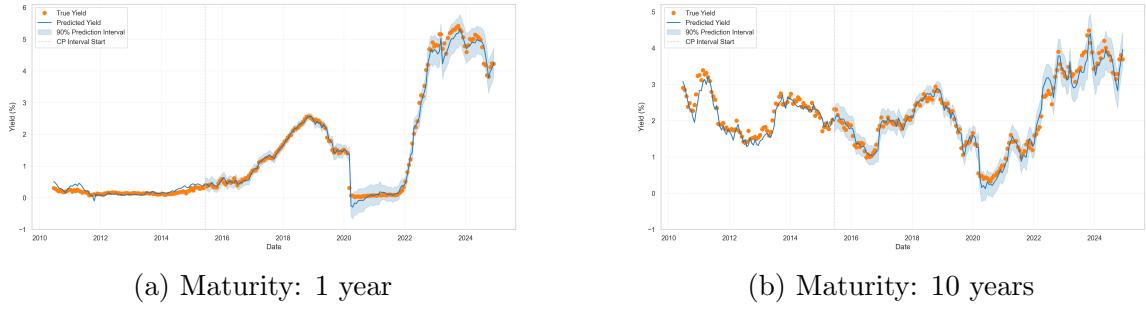


Figure 7.2: Two-step DNS VAR(1) forecast with conformal prediction intervals for federal funds. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .

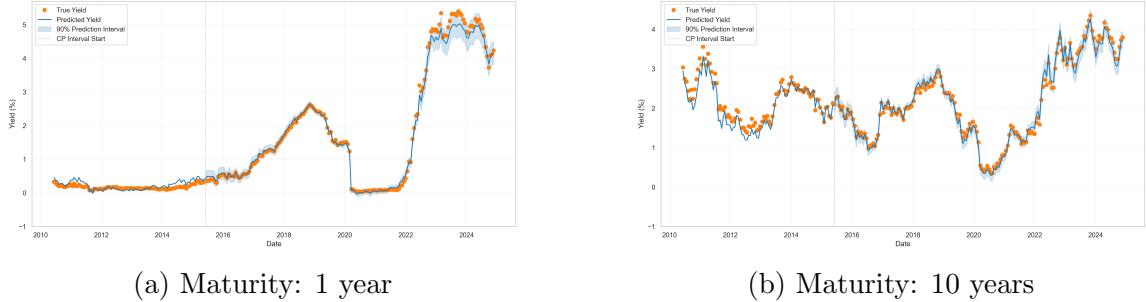


Figure 7.3: Bootstrap forecast with conformal prediction intervals for federal funds. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window, 1000 samples,  $L \in [8, 12]$  days,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .

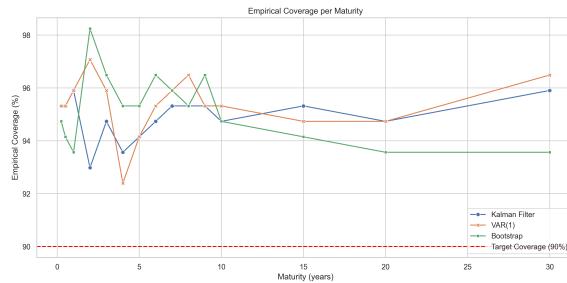


Figure 7.4: Empirical coverage of the conformal prediction interval width over time by maturity and model for federal funds, coverage level 90%

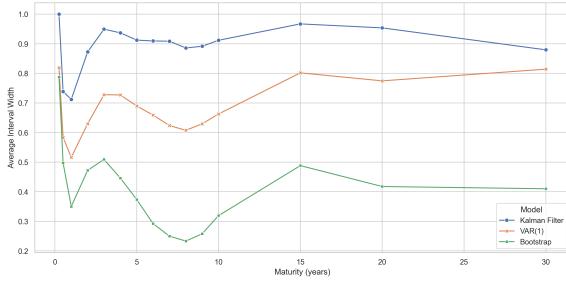


Figure 7.5: Average conformal prediction interval width over time by maturity and model for federal funds, coverage level 90%

The one-step DNS model with the Kalman filter produces noisy forecasts, with prediction intervals that widen considerably during periods of elevated market volatility, such as in 2020–2021 and again in 2023–2024. The intervals are the widest among the three models across all maturities. Despite the higher volatility in the predictions, the conformal intervals maintain a stable coverage level of around 95% across maturities.

The two-step DNS model with VAR(1) yields smoother forecasts and narrower intervals than the one-step DNS. The intervals widen during volatile periods, though to a lesser extent than in the one-step model. Interval width and coverage vary more across maturities, particularly at the short end, where the empirical coverage ranges between 92% and 97%.

The bootstrap model exhibits distinct behaviour compared to the 1- and 2-step DNS. The forecasts alternate between periods of under- and overshooting; the predicted yields often lie near the edges of the true yield range rather than at the centre, as observed in the other models. The conformal intervals are the narrowest overall but still achieve good empirical coverage, particularly at shorter maturities. Forecast behaviour appears more stable over time, even during volatile periods.

### 7.1.2 Mean Squared Error and Pearson Correlation

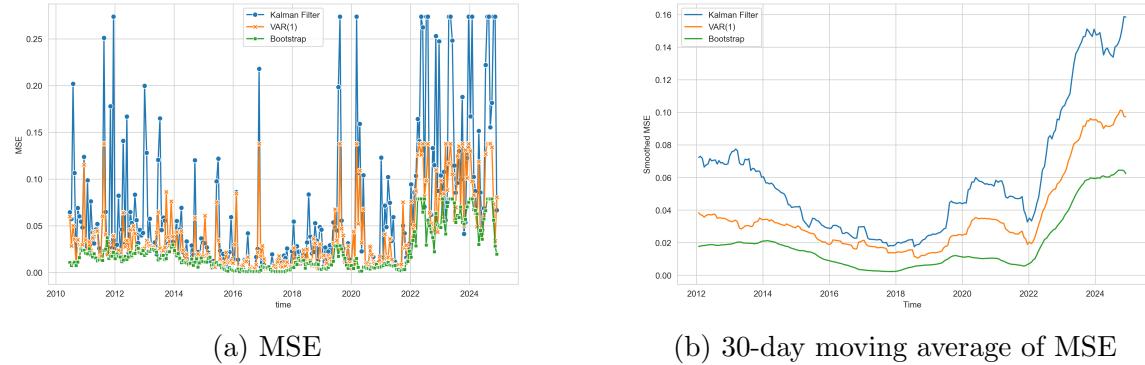


Figure 7.6: Mean Squared Error (MSE) over time for 14-day ahead forecasts of federal funds. Left: raw daily values. Right: 30-day moving average.

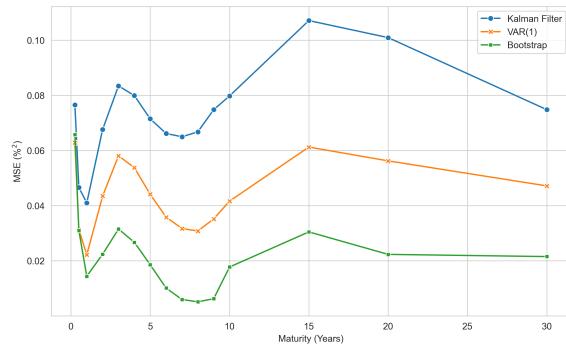


Figure 7.7: Mean Squared Error (MSE) by maturity for 14-day ahead forecasts of federal funds.

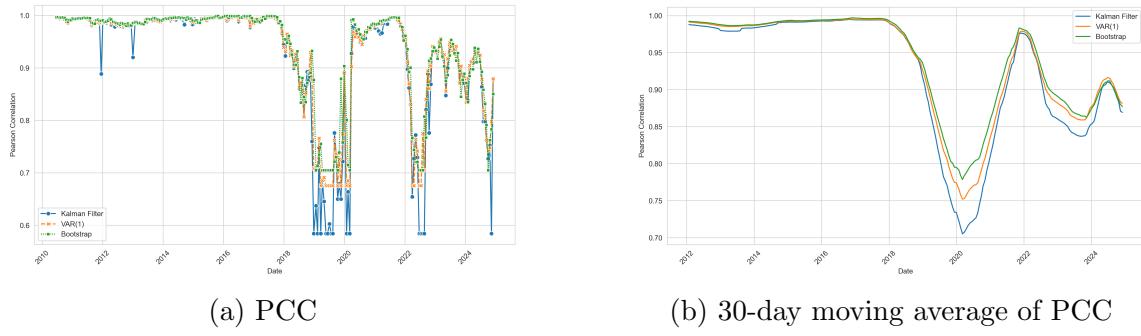


Figure 7.8: Pearson Correlation Coefficient (PCC) over time for 14-day ahead forecasts of federal funds. Left: raw daily values. Right: 30-day moving average.

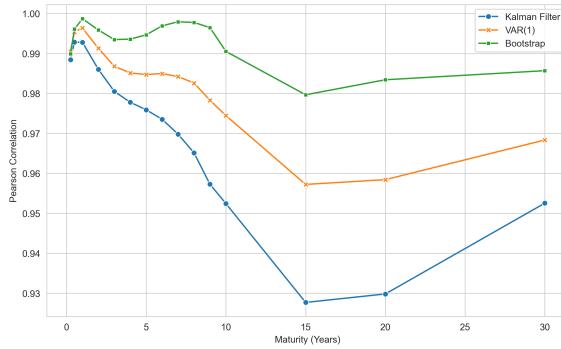


Figure 7.9: Pearson Correlation Coefficient (PCC) by maturity for 14-day ahead forecasts of federal funds.

Maturity	One-Step DNS		Two-Step DNS		Bootstrap	
	MSE	PCC	MSE	PCC	MSE	PCC
0.25	0.076530	0.988439	0.062790	0.990496	0.065724	0.989919
0.5	0.046585	0.992851	0.031289	0.995408	0.030947	0.996094
1	0.040993	0.992816	0.022159	0.996383	0.014377	0.998700
2	0.067634	0.986022	0.043459	0.991300	0.022382	0.995889
3	0.083423	0.980519	0.058048	0.986800	0.031495	0.993485
4	0.079873	0.977795	0.053751	0.985132	0.026679	0.993601
5	0.071450	0.975905	0.044093	0.984746	0.018537	0.994676
6	0.066148	0.973535	0.035828	0.984979	0.010098	0.996928
7	0.064957	0.969807	0.031684	0.984230	0.005954	0.997945
8	0.066732	0.965128	0.030788	0.982617	0.005106	0.997777
9	0.074780	0.957277	0.035180	0.978288	0.006227	0.996484
10	0.079761	0.952435	0.041623	0.974460	0.017733	0.990519
15	0.107088	0.927699	0.061215	0.957236	0.030457	0.979669
20	0.100906	0.929815	0.056216	0.958430	0.022307	0.983428
30	0.074791	0.952535	0.047142	0.968349	0.021516	0.985701

Table 7.1: Mean Squared Errors (MSE) and Pearson correlations between predicted and observed yields across maturities, for federal funds.

In terms of mean squared error (MSE), the bootstrap method consistently achieves the lowest error across time and maturities, see figures 7.6 and 7.7. In contrast, the one-step DNS model exhibits the highest and most variable MSE, particularly during periods of heightened volatility, such as 2022–2024. This pattern is apparent in the daily series and the 30-day moving average. The two-step DNS model lies between the two, with moderate and stable error levels over time. After 2022, the MSE increases sharply for all three methods. The error is the lowest right before, between 2018 and 2022. Across maturities, the one-step DNS show a notable increase in MSE around the 4-year and 15-year curves, while the two-step DNS and bootstrap models show slightly flatter profiles.

The Pearson correlation coefficient (PCC) supports this pattern, see figures 7.8 and 7.9. All three models achieve high PCC across maturities, but the bootstrap method demonstrates the strongest alignment with the observed yields, particularly at shorter

maturities. Over time, all models experience sharp drops in PCC during market stress periods, notably during 2019 and 2021. This behaviour is most prominent in the one-step DNS, reflecting its greater sensitivity to market shocks. The two-step DNS is more stable over time but exhibits weaker correlation at longer maturities. Over time, the bootstrap model performs similarly to the two-step DNS but performs much better across maturities.

### 7.1.3 Statistical Properties

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.6016	-0.2560	0.1169	0.1852
0.5	0.4409	-0.2413	0.1220	0.1749
1	0.2996	-0.2139	0.1326	0.1535
2	0.2106	-0.1750	0.1539	0.1136
3	0.1743	-0.1633	0.1730	0.0833
4	0.1691	-0.1727	0.1884	0.0626
5	0.1726	-0.1921	0.1998	0.0492
6	0.1658	-0.2133	0.2081	0.0405
7	0.1540	-0.2327	0.2140	0.0348
8	0.1490	-0.2490	0.2183	0.0311
9	0.1467	-0.2623	0.2215	0.0287
10	0.1624	-0.2729	0.2240	0.0271
15	0.1352	-0.2988	0.2306	0.0237
20	0.1275	-0.3013	0.2261	0.0208
30	0.1058	-0.3036	0.1316	0.0126

Table 7.2: Autocorrelation of 14-day yield differences for US federal funds, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.0165	0.0167	0.0160	0.0157
0.5	0.0161	0.0162	0.0155	0.0152
1	0.0151	0.0151	0.0145	0.0143
2	0.0131	0.0131	0.0127	0.0125
3	0.0110	0.0113	0.0111	0.0109
4	0.0091	0.0097	0.0096	0.0094
5	0.0073	0.0082	0.0083	0.0080
6	0.0063	0.0069	0.0071	0.0068
7	0.0054	0.0057	0.0060	0.0057
8	0.0047	0.0047	0.0050	0.0047
9	0.0044	0.0037	0.0042	0.0038
10	0.0030	0.0029	0.0034	0.0031
15	0.0017	0.0002	0.0008	0.0004
20	0.0007	-0.0008	-0.0002	-0.0007
30	-0.0008	0.0003	0.0005	-0.0000

Table 7.3: Mean of 14-day yield differences for US federal funds, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.0126	0.0722	0.0288	0.0198
0.5	0.0152	0.0685	0.0280	0.0198
1	0.0199	0.0630	0.0267	0.0201
2	0.0251	0.0579	0.0255	0.0217
3	0.0283	0.0590	0.0256	0.0243
4	0.0303	0.0642	0.0265	0.0275
5	0.0319	0.0721	0.0279	0.0310
6	0.0324	0.0815	0.0296	0.0345
7	0.0330	0.0915	0.0314	0.0378
8	0.0330	0.1013	0.0332	0.0409
9	0.0329	0.1106	0.0348	0.0436
10	0.0325	0.1189	0.0363	0.0459
15	0.0332	0.1417	0.0395	0.0506
20	0.0326	0.1364	0.0373	0.0466
30	0.0307	0.1046	0.0310	0.0314

Table 7.4: Variance of 14-day yield differences for US federal funds, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	-2.3986	-0.9439	-4.2129	-1.5191
0.5	-1.9621	-0.9172	-4.0858	-1.3447
1	-1.4275	-0.8528	-3.7754	-1.0161
2	-0.7144	-0.6755	-3.0121	-0.5079
3	-0.3895	-0.4481	-2.2221	-0.2172
4	-0.2595	-0.2102	-1.5436	-0.0802
5	-0.1708	0.0042	-1.0244	-0.0269
6	-0.1883	0.1823	-0.6530	-0.0140
7	-0.1717	0.3254	-0.3967	-0.0190
8	-0.1554	0.4395	-0.2229	-0.0299
9	-0.1516	0.5313	-0.1062	-0.0417
10	-0.0998	0.6060	-0.0288	-0.0527
15	-0.1955	0.8298	0.0668	-0.0844
20	-0.2357	0.9000	-0.1136	-0.1018
30	-0.3222	0.4075	-1.7637	-0.3956

Table 7.5: Skewness of 14-day yield differences for US federal funds, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	36.0220	13.4539	49.0835	27.9716
0.5	32.3798	13.5308	47.2454	25.5106
1	18.0506	13.2442	42.8693	20.6194
2	7.8054	10.8526	32.6906	12.3811
3	4.0205	7.3247	23.0204	7.0892
4	2.4352	4.5612	15.4869	4.1259
5	1.6634	3.1646	10.2836	2.5427
6	1.4645	2.7858	6.9172	1.6948
7	1.4124	2.9511	4.8014	1.2283
8	1.3575	3.3474	3.4808	0.9591
9	1.4421	3.8115	2.6538	0.7942
10	1.5034	4.2681	2.1334	0.6858
15	1.6488	5.8676	1.4645	0.4267
20	1.8219	6.3091	2.3663	0.3252
30	1.9533	4.2996	16.1559	2.2010

Table 7.6: Kurtosis of 14-day yield differences for US federal funds, comparing observed yields with forecasts.

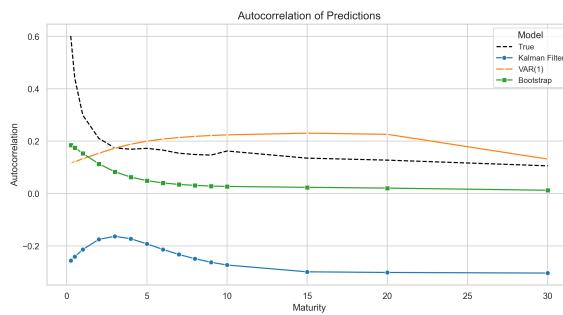


Figure 7.10: Autocorrelation of 14-day yield differences for US federal funds, averaged over time.

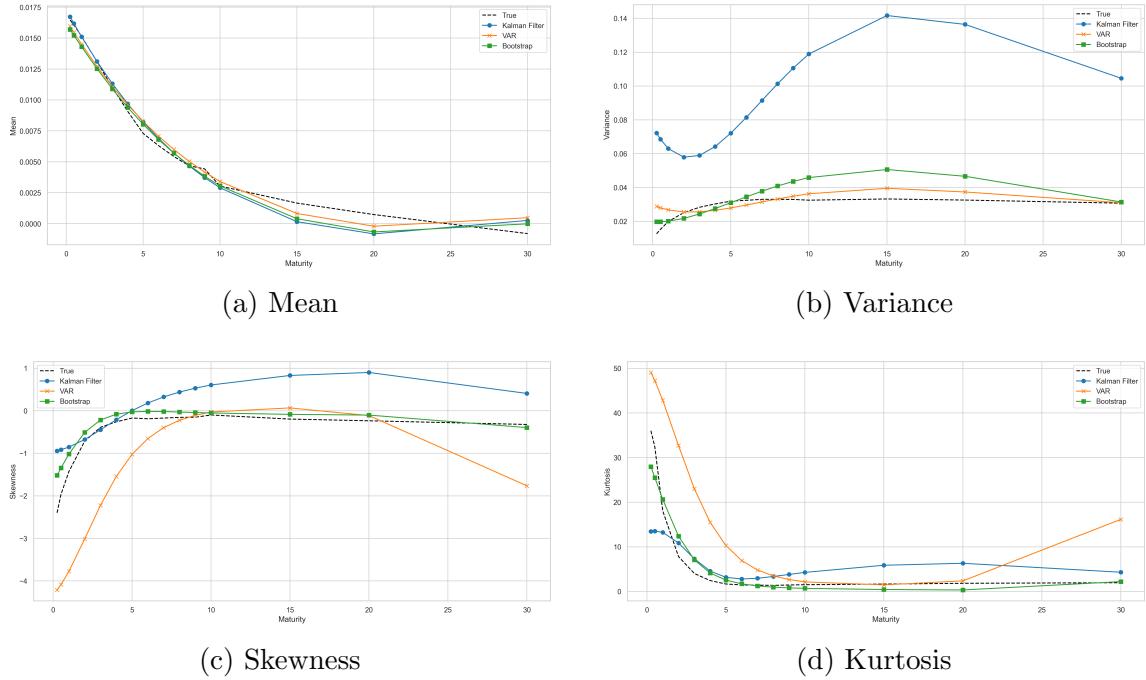


Figure 7.11: Distributional moments of 14-day yield differences for US federal funds, averaged over time.

Figure 7.10 presents the autocorrelation of 14-day yield differences across maturities for observed yields and forecasts produced by the three models. The observed series displays a decline in autocorrelation, starting at around 0.6 for the shortest maturities and decreasing to values between 0.1 and 0.2 beyond ten years. The bootstrap model captures the overall level reasonably well, although it consistently underestimates autocorrelation across most of the curve. The two-step DNS VAR(1) model tends to overestimate the autocorrelation beyond five years. The one-step DNS Kalman filter shows the greatest deviation from the observed pattern, producing negative autocorrelation estimates between -0.2 and -0.3 across all maturities.

Figure 7.11 shows the distributional moments of the 14-day yield differences. All models replicate the mean accurately. The variance is well captured by the bootstrap and two-step DNS models, while the one-step model overestimates it, with the largest deviations around the 15-year maturity. The bootstrap model closely match the skewness and kurtosis over the full maturity range. The one-step model approximates skewness reasonably well but overstates it slightly at both the short and long ends of the curve. For kurtosis, it aligns fairly well with the observed data, though with slight

underestimation at shorter maturities. The two-step model underestimates skewness up to the 10-year point, matches it between 10 and 20 years, and underestimates again toward the 30-year maturity. For kurtosis, the same maturity pattern appears, but with consistent overestimation instead.

## 7.2 Euro Interbank Offered Rate (Euribor)

### 7.2.1 Conformal Prediction

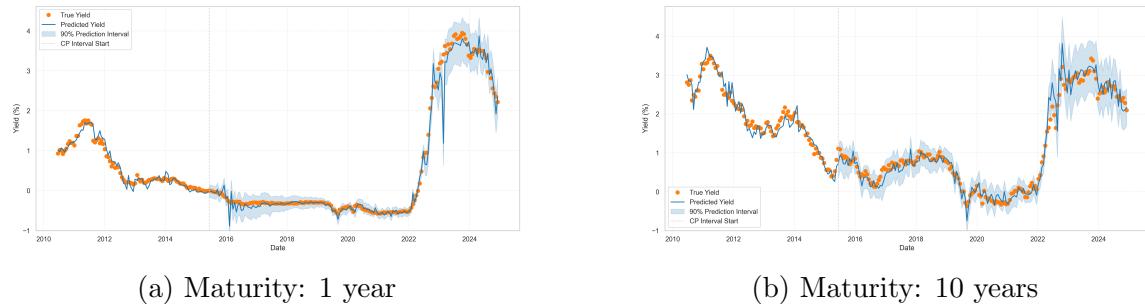


Figure 7.12: One-step DNS Kalman filter forecast with conformal prediction intervals for Euribor. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .

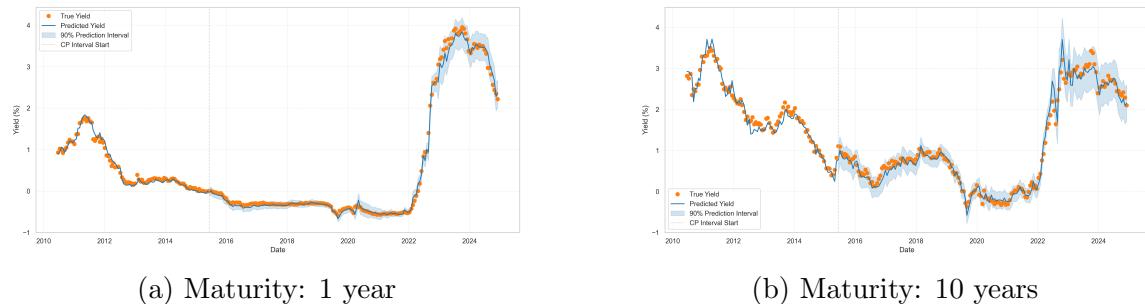


Figure 7.13: Two-step DNS VAR(1) forecast with conformal prediction intervals for Euribor. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .

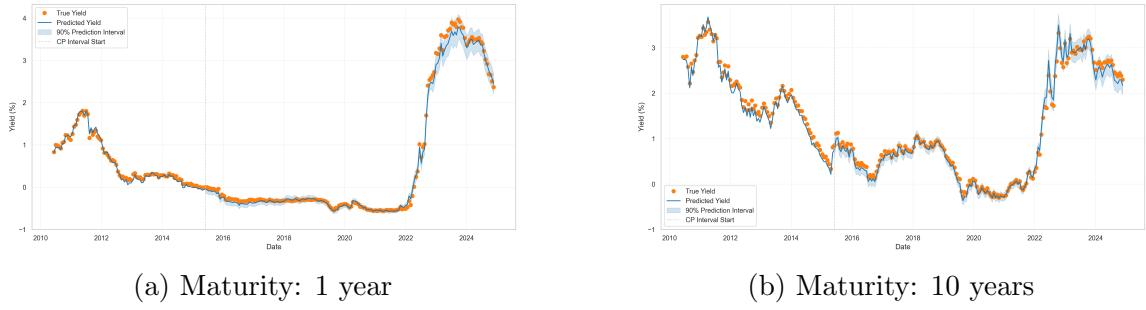


Figure 7.14: Bootstrap forecast with conformal prediction intervals for Euribor. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window, 1000 samples,  $L \in [8, 12]$  days,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$

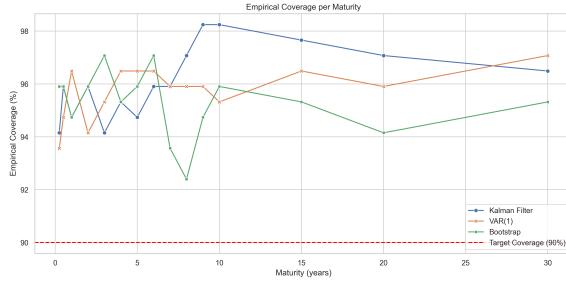


Figure 7.15: Empirical coverage of the conformal prediction interval width over time by maturity and model for Euribor, coverage level 90%

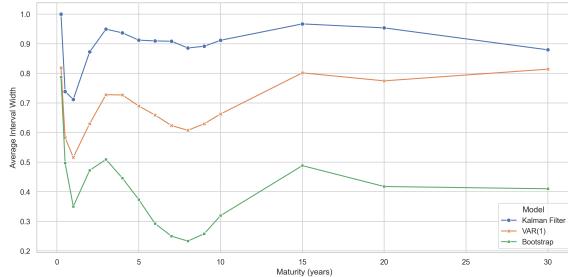


Figure 7.16: Average conformal prediction interval width over time by maturity and model for Euribor, coverage level 90%

The one-step DNS model with the Kalman filter produces noisy predictions that respond visibly to periods of increased volatility. Sharp spikes are observed around 2016 and 2023 for the 1-year curve, and around 2022-2023 for the 10-year curve. The prediction intervals widen significantly during periods of elevated uncertainty, particularly between 2016 and 2018 and from 2023 onward for the 1-year curve, and after 2022 for the 10-year curve. The prediction intervals are the widest among the three models across all maturities. The empirical coverage is high, consistently exceeding 94% and reaching over 98% for the 10-year maturity.

The two-step DNS model with VAR(1) results in considerably smoother predictions than the one-step model, with instability primarily present around 2022-2023 for the 10-year curve. The conformal intervals expand during periods of volatility, such as after 2022, but remain tight overall. The empirical coverage is stable and remains above the nominal level across all maturities, typically ranging between 94% and 97%.

The bootstrap model produces smooth forecasts that closely track the observed yields. As with the federal funds, the predicted values often lie near the upper or lower edge of the true yields, rather than in the centre. The intervals expand moderately during high-volatility periods but remain narrower overall than those from the 1- and two-step DNS models. The prediction intervals are tight but variable, particularly at short to medium maturities. A notable drop occurs around the 7-year curve, corresponding to a drop in empirical coverage to 92%. Aside from this dip, the empirical coverage is high.

### 7.2.2 Mean Squared Error and Pearson Correlation

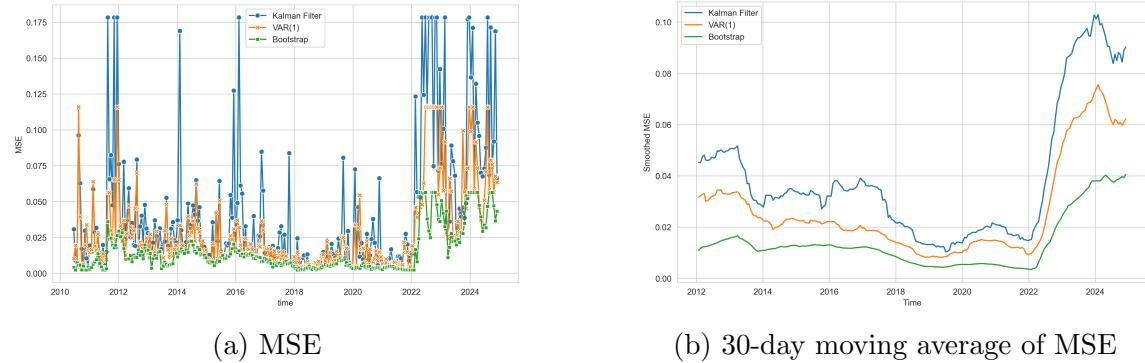


Figure 7.17: Mean Squared Error (MSE) over time for 14-day ahead forecasts of Euribor. Left: raw daily values. Right: 30-day moving average.

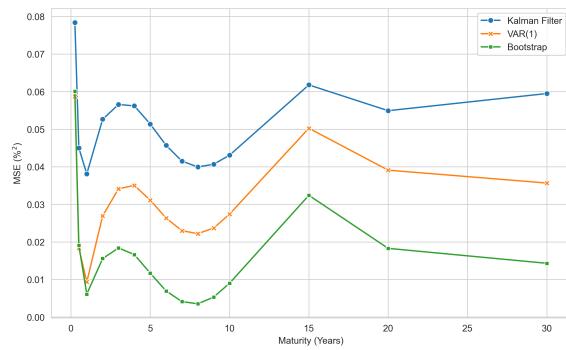


Figure 7.18: Mean Squared Error (MSE) by maturity for 14-day ahead forecasts of Euribor.

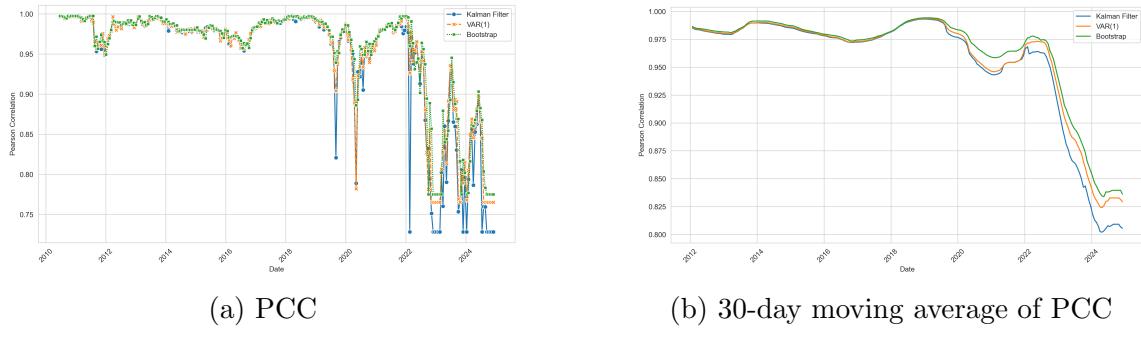


Figure 7.19: Pearson Correlation Coefficient (PCC) over time for 14-day ahead forecasts of Euribor. Left: raw daily values. Right: 30-day moving average

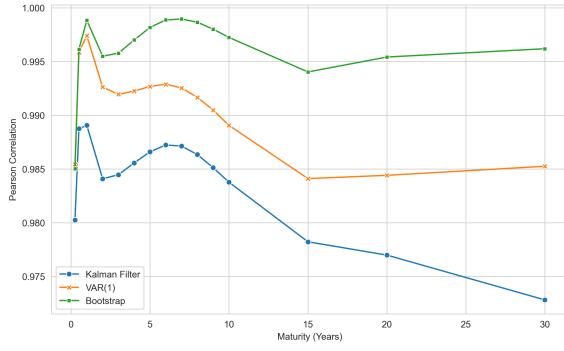


Figure 7.20: Pearson Correlation Coefficient (PCC) by maturity for 14-day ahead forecasts of Euribor.

Maturity	One-Step DNS		Two-Step DNS		Bootstrap	
	MSE	PCC	MSE	PCC	MSE	PCC
0.25	0.078371	0.980243	0.058701	0.985427	0.060054	0.985014
0.5	0.045008	0.988744	0.018398	0.995858	0.019062	0.996145
1	0.038083	0.989058	0.009322	0.997404	0.006082	0.998842
2	0.052627	0.984087	0.026917	0.992627	0.015557	0.995496
3	0.056570	0.984463	0.034126	0.991944	0.018378	0.995778
4	0.056194	0.985567	0.035035	0.992252	0.016622	0.997004
5	0.051369	0.986595	0.031108	0.992683	0.011678	0.998178
6	0.045700	0.987234	0.026336	0.992888	0.006930	0.998871
7	0.041485	0.987126	0.022976	0.992541	0.004108	0.998961
8	0.039942	0.986339	0.022190	0.991666	0.003512	0.998655
9	0.040664	0.985119	0.023670	0.990482	0.005308	0.998014
10	0.043076	0.983769	0.027391	0.989054	0.008986	0.997246
15	0.061776	0.978220	0.050239	0.984105	0.032392	0.994031
20	0.054916	0.976979	0.039087	0.984406	0.018276	0.995420
30	0.059481	0.972811	0.035660	0.985250	0.014297	0.996193

Table 7.7: Mean Squared Errors (MSE) and Pearson correlations between predicted and observed yields across maturities, for Euribor.

The bootstrap method has the lowest mean squared error (MSE) throughout the sample, both over time and across maturities (figures 7.17 and 7.18). The one-step DNS shows the highest MSE levels, with the two-step DNS performing between the two. These relative differences persist over time, with all three models experiencing a clear increase in MSE following the onset of increased market volatility in 2022. The MSE is the lowest right before, between 2018 and 2022.

The Pearson correlation coefficient (PCC) patterns shown in figures 7.19 and 7.20 align with the MSE results. All three models exhibit instability and sharp drops in correlation after 2022, with the one-step DNS being the most extreme, evident in figure 7.19a. Aside from this, their performance is similar. Across maturities, the bootstrap model maintains a high degree of correlation with observed yields; over 99.5%, except for the 3-month curve (figure 7.20). The one-step DNS exhibits the

weakest performance, dropping down to 97.5% for the 30-year curve. Nonetheless, all models maintain overall high correlation values.

### 7.2.3 Statistical Properties

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.5654	-0.3657	0.2398	0.2066
0.5	0.5444	-0.3532	0.2360	0.1896
1	0.4191	-0.3275	0.2217	0.1568
2	0.2424	-0.2789	0.1768	0.1014
3	0.1777	-0.2446	0.1282	0.0616
4	0.1487	-0.2318	0.0870	0.0348
5	0.1278	-0.2372	0.0550	0.0171
6	0.1277	-0.2526	0.0308	0.0049
7	0.1222	-0.2710	0.0126	-0.0036
8	0.1186	-0.2885	-0.0014	-0.0100
9	0.1164	-0.3034	-0.0124	-0.0151
10	0.1122	-0.3154	-0.0213	-0.0195
15	0.1010	-0.3402	-0.0494	-0.0388
20	0.0940	-0.3304	-0.0695	-0.0608
30	0.0539	-0.3101	-0.1175	-0.1021

Table 7.8: Autocorrelation of 14-day yield differences for Euribor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.0086	0.0063	0.0068	0.0074
0.5	0.0069	0.0059	0.0064	0.0070
1	0.0050	0.0052	0.0056	0.0062
2	0.0034	0.0038	0.0042	0.0048
3	0.0019	0.0025	0.0029	0.0036
4	0.0010	0.0014	0.0018	0.0024
5	0.0001	0.0004	0.0007	0.0014
6	-0.0007	-0.0005	-0.0002	0.0005
7	-0.0014	-0.0013	-0.0010	-0.0003
8	-0.0019	-0.0021	-0.0018	-0.0010
9	-0.0024	-0.0027	-0.0024	-0.0017
10	-0.0028	-0.0032	-0.0030	-0.0022
15	-0.0041	-0.0050	-0.0048	-0.0040
20	-0.0047	-0.0056	-0.0053	-0.0044
30	-0.0047	-0.0045	-0.0042	-0.0032

Table 7.9: Mean of 14-day yield differences for Euribor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.0068	0.0705	0.0132	0.0116
0.5	0.0076	0.0659	0.0131	0.0118
1	0.0103	0.0583	0.0131	0.0123
2	0.0145	0.0482	0.0143	0.0141
3	0.0163	0.0434	0.0164	0.0164
4	0.0176	0.0424	0.0192	0.0190
5	0.0188	0.0441	0.0222	0.0217
6	0.0193	0.0475	0.0253	0.0245
7	0.0198	0.0519	0.0283	0.0271
8	0.0204	0.0567	0.0311	0.0294
9	0.0210	0.0615	0.0336	0.0316
10	0.0216	0.0660	0.0358	0.0334
15	0.0235	0.0798	0.0413	0.0384
20	0.0232	0.0792	0.0395	0.0375
30	0.0217	0.0733	0.0304	0.0300

Table 7.10: Variance of 14-day yield differences for Euribor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	1.2102	1.5174	2.9851	1.6638
0.5	1.8244	1.5540	3.0682	1.6464
1	1.5133	1.6201	3.1709	1.5951
2	0.8271	1.7090	3.1222	1.4567
3	0.6505	1.7211	2.8506	1.3076
4	0.5237	1.6595	2.5117	1.1717
5	0.3888	1.5513	2.1912	1.0582
6	0.3636	1.4263	1.9174	0.9644
7	0.2786	1.3039	1.6927	0.8877
8	0.2005	1.1931	1.5108	0.8251
9	0.1789	1.0962	1.3637	0.7745
10	0.1226	1.0124	1.2445	0.7339
15	-0.0284	0.7288	0.9058	0.6203
20	-0.1590	0.5448	0.7701	0.5735
30	-0.2964	0.2718	0.4811	0.3350

Table 7.11: Skewness of 14-day yield differences for Euribor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	10.3802	30.7467	20.0493	10.8691
0.5	12.4310	29.8498	20.9764	10.9550
1	9.9064	27.4271	22.2815	10.8119
2	5.8670	20.8079	22.3876	9.6454
3	4.3454	14.2852	20.2053	8.0704
4	3.4193	10.2119	17.4014	6.6323
5	2.6806	8.5438	14.8536	5.4969
6	2.2503	8.0292	12.7896	4.6375
7	2.0472	7.7689	11.1767	3.9945
8	1.9446	7.4454	9.9196	3.5118
9	1.7760	7.0263	8.9277	3.1484
10	1.7818	6.5531	8.1299	2.8727
15	1.4543	4.2712	5.6620	2.1992
20	1.2903	2.4748	4.2283	2.0276
30	1.2011	1.5459	3.1969	1.9775

Table 7.12: Kurtosis of 14-day yield differences for Euribor, comparing observed yields with forecasts.

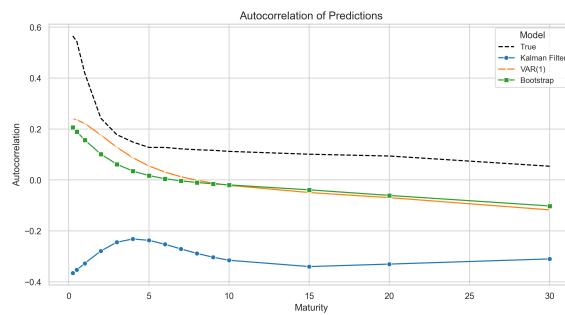


Figure 7.21: Autocorrelation of 14-day yield differences for Euribor, averaged over time.

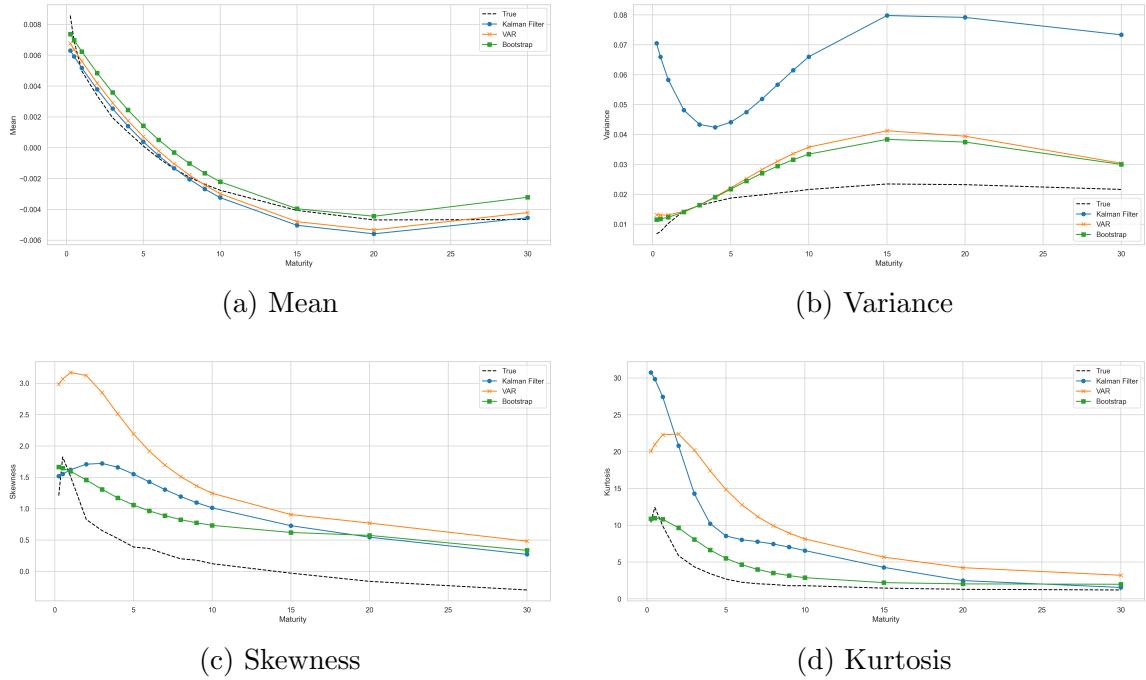


Figure 7.22: Distributional moments of 14-day yield differences for Euribor, averaged over time.

Figure 7.21 presents the autocorrelations of 14-day yield differences across maturities for observed Euribor yields and forecasts. All models underestimate the empirical autocorrelations to varying degrees. The two-step DNS and bootstrap models perform similarly and provide the most accurate approximations, capturing both the general level and shape of the observed series. In contrast, the one-step DNS Kalman filter underestimates the autocorrelations more substantially and fails to reproduce the overall pattern observed in the data.

Figure 7.22 displays the first four distributional moments of the 14-day yield differences. All models reproduce the mean with high accuracy. For variance, the bootstrap and two-step DNS models yield nearly identical results. These estimates closely match the observed variance for maturities below five years but tend to overestimate it for longer maturities. The one-step model deviates more significantly, overestimating both the level and the shape of the variance curve. All three models approximate the general trend in skewness but consistently overestimate its magnitude. The two-step model exhibits the largest and most consistent overestimation. For kurtosis, the bootstrap model provides the closest match to the observed data

across the curve. Both the one-step and two-step models overestimate kurtosis, particularly at shorter maturities.

## 7.3 Stockholm Interbank Offered Rate (Stibor)

### 7.3.1 Conformal Prediction

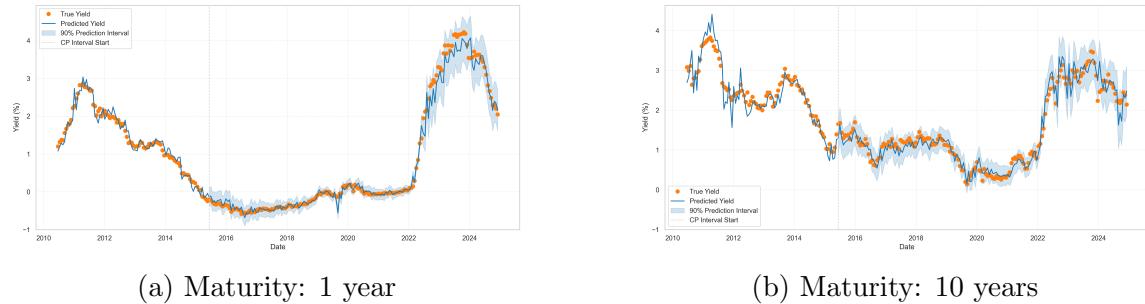


Figure 7.23: One-step DNS Kalman filter forecast with conformal prediction intervals for Stibor. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .

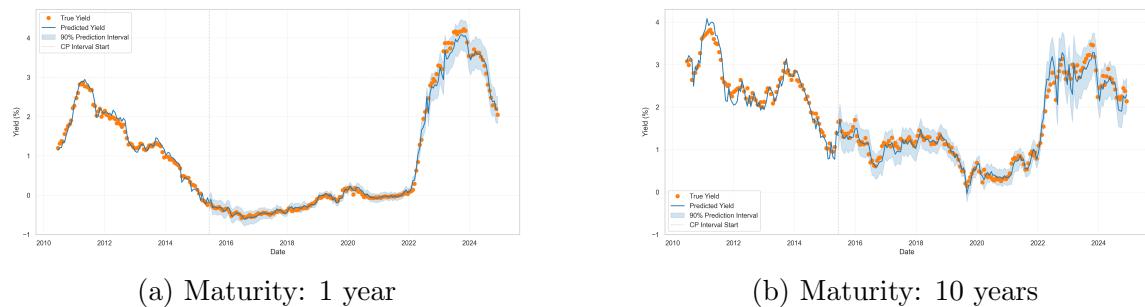
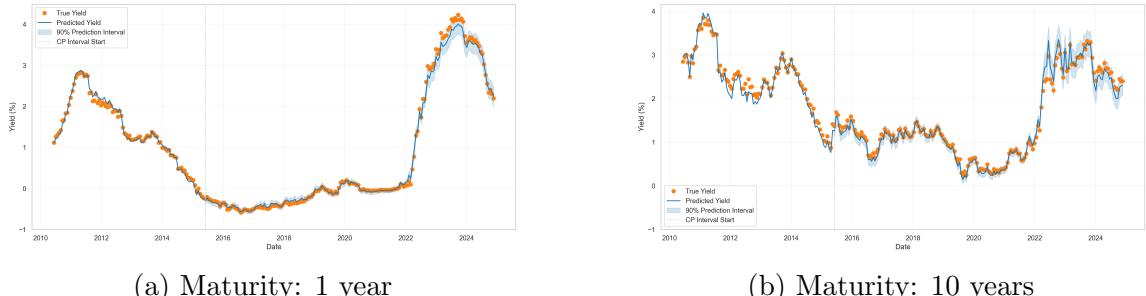


Figure 7.24: Two-step DNS VAR(1) forecast with conformal prediction intervals for Stibor. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .



(a) Maturity: 1 year

(b) Maturity: 10 years

Figure 7.25: Bootstrap forecast with conformal prediction intervals for Stibor. Forecasted using daily data, 14-day forecast-horizon, 360-day rolling window, 1000 samples,  $L \in [8, 12]$  days,  $\alpha = 0.0609$ . Conformal prediction calibration window: 5 years, coverage level 90%, weight decay parameter  $\rho = 0.96$ .

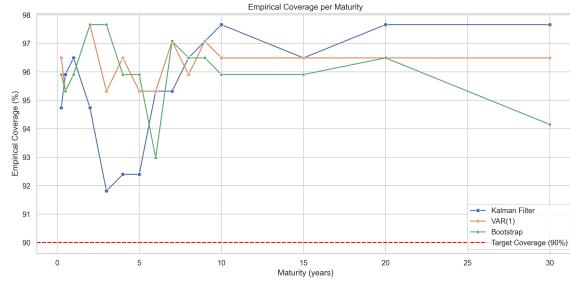


Figure 7.26: Empirical coverage of the conformal prediction interval width over time by maturity and model for Stibor, coverage level 90%

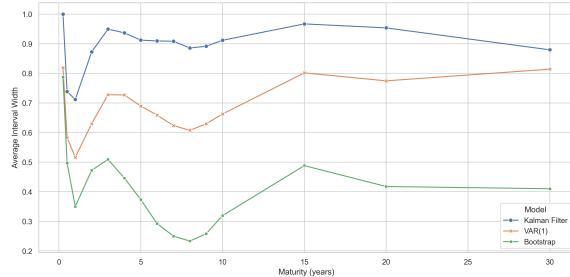


Figure 7.27: Average conformal prediction interval width over time by maturity and model for Stibor, coverage level 90%

The one-step DNS model with the Kalman filter produces noisy forecasts (figure 7.23). At the 1-year maturity, fluctuations are especially noticeable from 2022 onward, accompanied by wide prediction intervals. At the 10-year maturity, the yield series is more volatile, with persistently wide prediction intervals. The prediction intervals widen substantially during 2022–2024, when market volatility is elevated. Among the three models, the Kalman filter produces the broadest conformal intervals across all maturities (figure 7.27). Empirical coverage remains above the nominal 90% level across the curve, but is variable and dips to around 92%–93% between years 3 and 5 (figure 7.26).

The two-step DNS model with VAR(1) produces smoother forecasts than the one-step version, with prediction intervals that widen in response to volatility but remain narrower overall (figure 7.24). The conformal intervals are moderately wide and show limited variation over time (figure 7.27). Empirical coverage is high and consistent across all maturities, remaining close to or above 96% (figure 7.26).

The bootstrap model produces the smoothest forecasts, and the predicted yields mostly follow the observed data closely (figure 7.25). As with previous cases, the forecasted yields tend to lie near the boundaries of the observed range rather than at the centre, although the tendency is less pronounced here. The conformal intervals are the narrowest of the three methods across all maturities (figure 7.27). Empirical coverage is generally stable and above the nominal threshold, with a slight decline at the 6- and 30-year maturities, where it drops to around 93% and 94% respectively (figure 7.26).

### 7.3.2 Mean Squared Error and Pearson Correlation

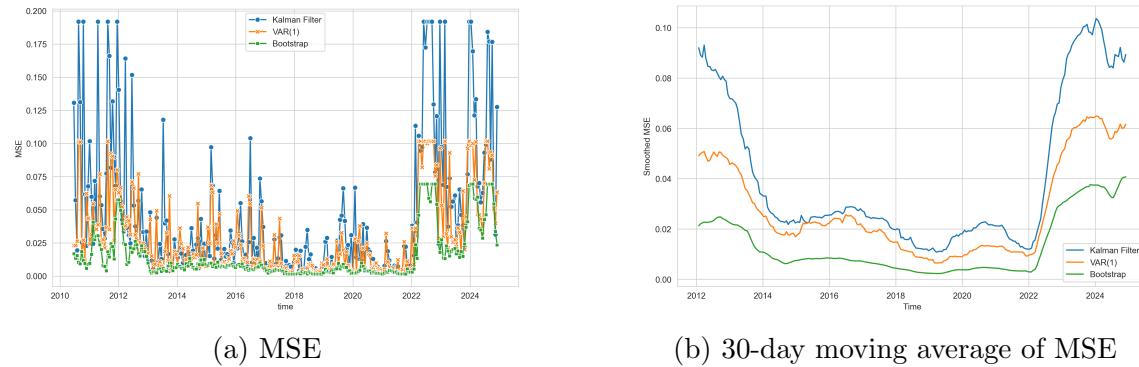


Figure 7.28: Mean Squared Error (MSE) over time for 14-day ahead forecasts of Stibor. Left: raw daily values. Right: 30-day moving average.

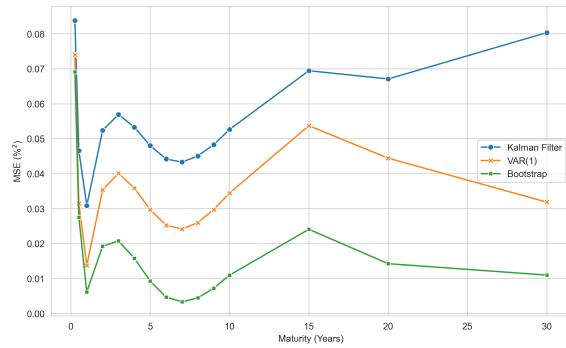


Figure 7.29: Mean Squared Error (MSE) per maturity for 14-day ahead forecasts of Stibor.

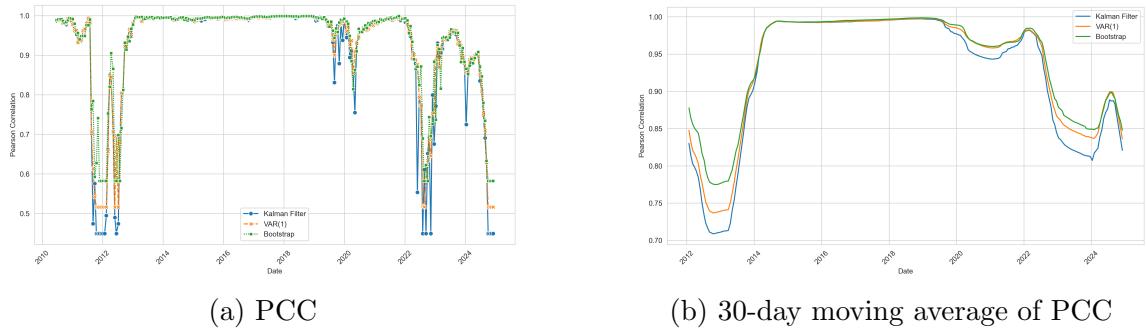


Figure 7.30: Pearson Correlation Coefficient (PCC) over time for 14-day ahead forecasts of Stibor. Left: raw daily values. Right: 30-day moving average.

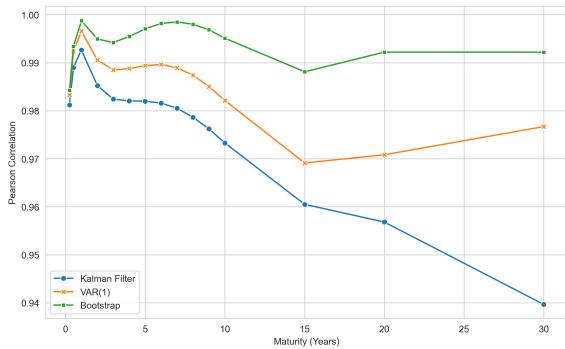


Figure 7.31: Pearson Correlation Coefficient per maturity for 14-day ahead forecasts of Stibor.

Maturity	One-Step DNS		Two-Step DNS		Bootstrap	
	MSE	PCC	MSE	PCC	MSE	PCC
0.25	0.083771	0.981205	0.074018	0.983257	0.069091	0.984274
0.5	0.046541	0.988994	0.031474	0.992353	0.027493	0.993395
1	0.030849	0.992632	0.013779	0.996599	0.006153	0.998742
2	0.052391	0.985255	0.035380	0.990559	0.019260	0.994936
3	0.056927	0.982413	0.040088	0.988483	0.020758	0.994210
4	0.053251	0.982059	0.035856	0.988796	0.015770	0.995454
5	0.048018	0.981985	0.029726	0.989400	0.009263	0.997053
6	0.044200	0.981605	0.025215	0.989637	0.004674	0.998194
7	0.043287	0.980520	0.024164	0.988966	0.003333	0.998491
8	0.045055	0.978624	0.025961	0.987401	0.004470	0.998008
9	0.048272	0.976221	0.029730	0.985029	0.007210	0.996851
10	0.052619	0.973287	0.034453	0.982155	0.010962	0.995105
15	0.069430	0.960510	0.053704	0.969116	0.024075	0.988112
20	0.067073	0.956811	0.044423	0.970831	0.014242	0.992206
30	0.080327	0.939664	0.031849	0.976696	0.010981	0.992197

Table 7.13: Mean Squared Errors (MSE) and Pearson correlations between predicted and observed yields across maturities, for Stibor.

The bootstrap model achieves the lowest mean squared error (MSE) for the Stibor forecasts, consistently outperforming both the 1- and 2-step DNS across the full sample period and maturities (figures 7.28 and 7.29). The one-step DNS model exhibits the highest MSE, with especially high errors during 2012–2014 and 2022–2024. The two-step DNS model lies between the two, yielding moderate and relatively stable error levels. Across maturities, the one-step DNS produces the highest errors, especially at the 3-month and 15–30 year maturities. At the same time, the two-step DNS remains intermediate, and the bootstrap model consistently delivers the lowest errors across the curve.

All three models achieve high Pearson correlation coefficients (PCC) over time, although all experience sharp drops in performance around 2012, 2022, and 2024, see figures 7.30 and 7.31. The one-step DNS displays the deeper declines in correlation, whereas the bootstrap model maintains the highest and most stable correlation

across time. When evaluated by maturity, the bootstrap model also outperforms the other two, with PCC values consistently exceeding 98% across the curve. On the contrary, the 1-step DNS shows a steady decline in correlation at longer maturities, reaching below 94% at the 30-year mark. The bootstrap approach offers the most stable and accurate alignment with actual yields in both dimensions.

### 7.3.3 Statistical Properties

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.4919	-0.3294	0.1064	0.2614
0.5	0.5684	-0.3265	0.0960	0.2390
1	0.3541	-0.3216	0.0751	0.1928
2	0.1959	-0.3165	0.0376	0.1111
3	0.1642	-0.3181	0.0108	0.0535
4	0.1480	-0.3236	-0.0060	0.0168
5	0.1335	-0.3299	-0.0158	-0.0056
6	0.1241	-0.3355	-0.0213	-0.0191
7	0.1178	-0.3401	-0.0244	-0.0273
8	0.1075	-0.3435	-0.0262	-0.0323
9	0.1010	-0.3461	-0.0273	-0.0354
10	0.0950	-0.3480	-0.0281	-0.0373
15	0.0756	-0.3539	-0.0329	-0.0396
20	0.0505	-0.3629	-0.0427	-0.0386
30	0.0351	-0.4091	-0.0748	-0.0234

Table 7.14: Autocorrelation of 14-day yield differences for Stibor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.0072	0.0048	0.0049	0.0052
0.5	0.0061	0.0046	0.0046	0.0049
1	0.0033	0.0041	0.0040	0.0043
2	0.0008	0.0034	0.0028	0.0032
3	-0.0004	0.0026	0.0017	0.0022
4	-0.0012	0.0020	0.0008	0.0012
5	-0.0018	0.0014	-0.0000	0.0004
6	-0.0023	0.0008	-0.0008	-0.0003
7	-0.0028	0.0004	-0.0015	-0.0010
8	-0.0031	-0.0001	-0.0021	-0.0016
9	-0.0034	-0.0005	-0.0026	-0.0021
10	-0.0036	-0.0008	-0.0030	-0.0025
15	-0.0042	-0.0020	-0.0045	-0.0040
20	-0.0052	-0.0025	-0.0049	-0.0044
30	-0.0059	-0.0024	-0.0038	-0.0035

Table 7.15: Mean of 14-day yield differences for Stibor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.0092	0.0499	0.0181	0.0115
0.5	0.0086	0.0472	0.0179	0.0116
1	0.0122	0.0431	0.0179	0.0121
2	0.0187	0.0393	0.0191	0.0139
3	0.0213	0.0399	0.0214	0.0165
4	0.0230	0.0436	0.0245	0.0197
5	0.0241	0.0492	0.0280	0.0231
6	0.0245	0.0559	0.0316	0.0266
7	0.0245	0.0631	0.0351	0.0300
8	0.0247	0.0702	0.0383	0.0331
9	0.0250	0.0769	0.0412	0.0358
10	0.0250	0.0830	0.0438	0.0382
15	0.0262	0.1026	0.0496	0.0439
20	0.0271	0.1085	0.0459	0.0410
30	0.0265	0.1304	0.0287	0.0253

Table 7.16: Variance of 14-day yield differences for Stibor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	0.9103	0.9773	1.2811	1.1706
0.5	1.0024	0.9438	1.1954	1.1301
1	0.8718	0.8543	1.0661	1.0329
2	0.3664	0.6282	0.9512	0.8125
3	0.1761	0.4319	0.9348	0.6098
4	0.0590	0.3165	0.9323	0.4508
5	-0.0358	0.2657	0.9149	0.3338
6	-0.1076	0.2493	0.8825	0.2483
7	-0.1513	0.2481	0.8415	0.1849
8	-0.1737	0.2534	0.7978	0.1387
9	-0.2056	0.2617	0.7551	0.1050
10	-0.2129	0.2722	0.7154	0.0804
15	-0.2398	0.3529	0.5759	0.0418
20	-0.2659	0.4493	0.5049	0.0821
30	-0.2234	0.1539	0.3072	0.2495

Table 7.17: Skewness of 14-day yield differences for Stibor, comparing observed yields with forecasts.

Maturity	True	One-Step DNS	Two-Step DNS	Bootstrap
0.25	3.3439	6.4101	8.2201	5.0753
0.5	3.8006	6.2569	7.5383	4.7669
1	4.7192	5.8358	6.5205	4.2645
2	4.3824	4.7514	5.5413	3.7309
3	3.5010	3.8397	5.1731	3.5240
4	2.3122	3.3655	4.8904	3.3974
5	1.6707	3.2198	4.5939	3.2828
6	1.3402	3.2348	4.3067	3.1709
7	1.1734	3.3112	4.0518	3.0646
8	1.0710	3.4048	3.8350	2.9652
9	1.0434	3.4976	3.6528	2.8752
10	1.0072	3.5819	3.4987	2.7941
15	0.9881	3.7772	2.9556	2.4934
20	1.1100	4.0844	2.5818	2.2751
30	1.0782	21.7303	2.5527	1.7647

Table 7.18: Kurtosis of 14-day yield differences for Stibor, comparing observed yields with forecasts.

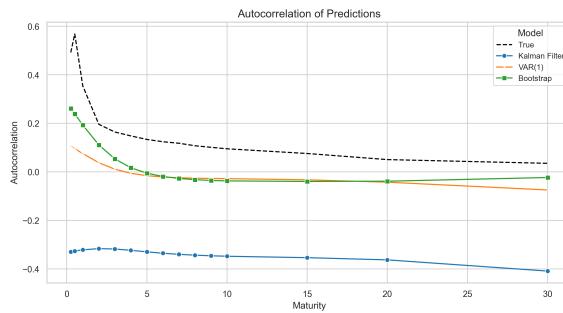


Figure 7.32: Autocorrelation of 14-day yield differences for Stibor, averaged over time.

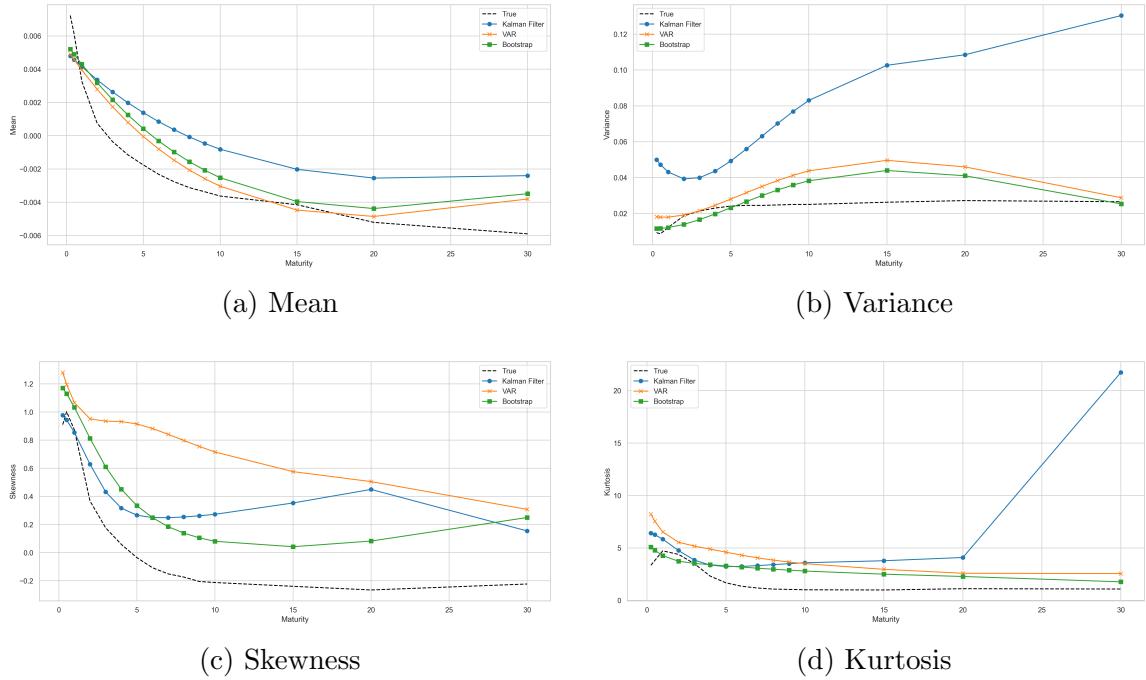


Figure 7.33: Distributional moments of 14-day yield differences for Stibor, averaged over time.

Figure 7.32 presents the autocorrelations of 14-day yield differences across maturities for observed Stibor yields and forecasts. All models underestimate the empirical autocorrelations to some extent. The two-step DNS and bootstrap models perform similarly and offer the most accurate approximations, capturing both the general level and overall shape of the observed series. The one-step DNS Kalman filter underestimates the autocorrelations more substantially and fails to replicate the observed pattern.

Figure 7.33 displays the first four distributional moments of the 14-day yield differences. While all models roughly capture the mean, the fit is less precise than in the federal funds and Euribor cases. The bootstrap and two-step models perform slightly better than the one-step model. Variance is reasonably well captured by the bootstrap and two-step models, though both slightly overestimate the empirical values around the 15-year maturity. The one-step model substantially overestimates the variance, particularly at longer maturities. All models overestimate the empirical skewness. The bootstrap model produces the closest match and also replicates the overall shape, whereas the two-step model exhibits the largest overestimation. Kur-

rosis is generally well captured across models, with the exception of a pronounced spike at the 30-year maturity in the one-step model, which is not reflected in the observed data.

## 7.4 Stability Analysis

This section evaluates the robustness of the forecasting results concerning key parameter choices made during model implementation. Rather than performing a grid search, which would be computationally intensive and less meaningful given the noisy, non-stationary nature of interest rate data, this analysis compares model performance across a few plausible parameter values. The goal is to assess whether the main findings remain stable when assumptions are varied within reasonable bounds.

### 7.4.1 One-Step DNS with Kalman Filter

The Kalman filter relies on several assumptions, including initialising the state-space matrices. While these significantly influence model behaviour, they are not user-defined hyperparameters in the usual sense. The most direct parameter under the user's control is the rolling window size used in the forecast.

The model was re-estimated to assess stability using two lengths of the rolling windows: 360 days and 1080 days. The changes in performance were minor. Some metrics, such as the Pearson correlation coefficient (PCC), slightly improved with longer windows, while others, like mean squared error (MSE), worsened marginally. This suggests that the one-step DNS is reasonably robust in the choice of estimation window. However, it is worth noting that the Kalman filter is known to be sensitive to initial values and model specification assumptions.

### 7.4.2 Two-Step DNS with VAR(1)

The two-step DNS model was also estimated using various rolling window lengths: 360 days and 1080 days. As with the one-step DNS, results remained relatively stable. While minor differences in performance metrics appeared across different window sizes, the overall forecast quality and statistical properties were preserved. This indicates that the two-step DNS method is reasonably robust concerning the estimation window.

### 7.4.3 Bootstrap DNS

The number of bootstrapped samples had a substantial impact on performance. With only 300 simulations, forecasts were unstable, MSE increased significantly, and the PCC often fell below 0.5, indicating weak forecast reliability. Increasing the number of samples to 1000 greatly improved stability and statistical performance. As expected, exponential smoothing was more noticeable when fewer samples were used. For larger sample sizes, the effect of smoothing became negligible.

Varying the block size had a relatively small effect on performance. Slightly better results were observed with smaller blocks, likely because they preserved finer temporal structure in the bootstrapped shocks. However, overall model behaviour remained stable across reasonable choices. This is likely because with a 14-day-ahead forecast horizon, all blocks are truncated to 14 days, reducing the impact of block size.

Changing the historical window length used for resampling (90, 180, 360 days) had little effect, likely because the sampling process was weighted toward more recent changes via exponential weights. This suggests that the method is robust to this parameter as long as the window is long enough to contain representative dynamics.

### 7.4.4 Conformal Prediction

This study uses a non-exchangeable split conformal prediction method, in which calibration residuals are weighted using a geometric decay factor  $\rho \in (0, 1]$ . When  $\rho = 1$ , the weights are uniform, corresponding to the standard exchangeable case.

The stability of the conformal intervals depends heavily on the choice of  $\rho$ . If  $\rho$  was set too low, the intervals became too narrow and the empirical coverage dropped below the nominal level. Conversely, setting  $\rho$  too close to 1 resulted in overly wide intervals approaching full coverage, indicating too-conservative intervals.

In comparison, the method was largely unaffected by the length of the calibration window, provided it contained a reasonable amount of data. Intervals computed from three, five or ten years of data were similar in quality, with only minor variations in width and coverage. This could be explained by the exponential weighting on recent data points.

# Chapter 8

## Discussion

### 8.1 Results Summary

This study highlights the strengths and weaknesses of three ways of forecasting yield curves within the Dynamic Nelson-Siegel model. While all methods produced reasonable forecasts during stable market conditions, the bootstrap-based DNS approach consistently outperformed the parametric alternatives across most evaluation criteria.

#### 8.1.1 Mean Squared Error and Pearson Correlation

The bootstrap method consistently achieved the lowest mean squared errors across time and maturities, and error levels were relatively stable over time. The one-step DNS (Kalman filter) yielded the highest errors, particularly during volatile periods, and error levels tended to be very noisy. The two-step DNS (VAR(1)) performed intermediately, with lower and more stable error levels than the one-step DNS. All models saw forecasting errors rise during market stress episodes, especially around 2022–2024.

The Pearson correlation between forecasts and actual yields was high for all methods in calm periods and across most maturities, but dropped sharply during periods of market stress. The one-step DNS had the sharpest declines in correlation — at times the PCC for the 14-day forecasts fell below 0.50 for Stibor 7.30, 0.75 for Euribor, and 0.6 for federal funds — indicating greater sensitivity to abrupt rate changes. While all three models saw significant drops in correlation during these times, the

bootstrap method maintained the strongest correlation with observed yields over time and across maturities.

### 8.1.2 Conformal Prediction Intervals

All three models produced conformal prediction intervals with empirical coverage above the nominal 90% level, consistent with the theoretical guarantees of conformal prediction. However, the width and dynamics of the prediction intervals varied slightly between the models.

The one-step DNS produced the widest intervals, particularly at longer maturities and during turbulent periods. The two-step DNS yielded narrower and smoother intervals, reflecting the more stable nature of the VAR(1) structure. The bootstrap method produced the narrowest intervals while still maintaining reliable coverage. Its intervals were also more stable over time, even during the volatile 2022–2024 period. This suggests that the bootstrap’s distribution-free simulation is more robust when capturing forecast uncertainty, particularly when market conditions shift rapidly.

### 8.1.3 Statistical Properties

In addition to assessing point forecasts and interval coverage, the models were evaluated based on their ability to preserve key statistical properties of the yield curve dynamics. Specifically, the autocorrelations and the first four distributional moments of 14-day yield differences were compared for both observed and predicted yields.

The observed autocorrelations displayed a downward-sloping profile characteristic of yield dynamics, where short maturities often exhibit higher persistence while long maturities tend to be more stable and less autocorrelated. The models underestimated the empirical autocorrelations to varying degrees for all three interest rate curves. The bootstrap model performed best and most closely replicated the observed autocorrelation structure. The two-step DNS model also produced reasonable estimates, particularly for Euribor and Stibor. The one-step DNS Kalman filter deviated substantially from the empirical pattern, exhibiting negative autocorrelation across the full maturity spectrum.

All models reproduced the mean of the yield changes well, although Stibor forecasts were slightly less precise. The bootstrap and two-step DNS models tracked the variance curve closely at short maturities but tended to overestimate it beyond five years. The one-step model significantly overestimated variance across the curve, likely due to the noisier yield estimates mentioned earlier.

The skewness was more difficult for the models to replicate. The bootstrap model captured it reasonably well, particularly for federal funds, although it slightly overestimated the level. The one-step model followed the general shape but consistently overestimated skewness across all maturities. The two-step model performed the worst, with large overestimations for Euribor and Stibor and underestimations for much of the federal funds curve.

The bootstrap model most accurately captured kurtosis, following the empirical pattern across all rates. The one-step model generally performed well but overestimated kurtosis for the short end, particularly for Euribor. It also produced an unrealistic spike at the 30-year maturity in the Stibor data. The two-step model consistently overestimated kurtosis, although its performance for Stibor was relatively close to the observed values.

#### 8.1.4 Interpretation of Results

The MSE and PCC results show that all three DNS-based models can capture yield dynamics well in normal conditions, while all struggle during market stress. The bootstrap method demonstrated the greatest resilience in turbulent environments, whereas the one-step DNS was more vulnerable to market stress and produced less stable forecasts with greater errors. The bootstrap method also performed the most consistently across the maturity spectrum.

As expected, the conformal prediction provided coverage well above the nominal 90% for all models. The bootstrap model offered the narrowest and most stable intervals, including during volatile periods. These results support the theoretical work by Vovk et al. (2005), showing that valid predictive intervals can be constructed without strong distributional assumptions. This is especially useful when comparing models that rely on different underlying assumptions.

It is worth noting that all models generated highly conservative intervals, with empirical coverage often reaching approximately 96%, well above the nominal 90%. While this conservatism is preferable to under-coverage from a risk management perspective, narrowing the intervals slightly to align with the nominal level could improve utility.

The bootstrap model most effectively replicated the autocorrelations and distributional moments. The difference was especially notable in the skewness and kurtosis. The one-step method struggled to replicate the autocorrelation and variance, while the two-step method had the worst approximations of skewness and kurtosis.

The observed differences in performance can be linked to the models' structural assumptions. The bootstrap method uses a non-parametric simulation approach, generating future paths by resampling historical changes. This could enable it to preserve the empirical distribution and serial dependencies without imposing rigid assumptions. In contrast, both the one- and two-step DNS models are parametric. The Kalman filter estimates the DNS system as a linear-Gaussian state-space model. While efficient under correct assumptions, this structure smooths shocks and limits the representation of fat tails and persistence. Similarly, the VAR(1) model in the two-step DNS assumes normally distributed innovations and linear factor dynamics, which may limit its ability to capture non-linearities in yield changes. The bootstrap method avoids such constraints, which could make it better suited for replicating complex empirical features.

## 8.2 Implications for Risk

The findings of this study have important implications for risk management, especially in areas such as market risk estimation, scenario generation, and stress testing. Although all three DNS-based models performed well under normal market conditions, their effectiveness in risk-related applications differs significantly.

### 8.2.1 Model Comparison

Risk managers are primarily concerned with the extremes of the yield distribution, the *tails*, rather than average yield changes. Inaccurate estimates of tail risk may lead to inappropriate capital buffers or misjudged exposure in stress scenarios. Skewness and kurtosis describe these tails; therefore, a model's ability to capture their empirical values is highly relevant. The bootstrap-based DNS model was most successful as it provided reasonable approximations across all rates and maturities. In contrast, the one- and two-step models tended to either over- or underestimate these moments.

Since financial exposures typically span multiple maturities, models must replicate yield changes across the entire curve, not only at individual maturities. The bootstrap model performed well across the maturity spectrum for all metrics, including MSE, PCC, autocorrelation, and distributional moments. The two-step DNS also performed reasonably well, although not as consistently. In contrast, the one-step model displayed more variation across maturities and struggled particularly with short-term dynamics.

Risk assessment requires predicting yield changes over a wide range of time horizons,

from one day to more than a year. Although long-term forecast accuracy cannot be directly validated through back-testing, models that reproduce the statistical properties and temporal dynamics of yield changes are more likely to generate realistic long-horizon simulations. The bootstrap model demonstrated strong performance in this regard, suggesting that it could be better suited for long-horizon stress simulations.

Since risk exposures typically span multiple asset classes and risk factors, the ability to capture dependencies among them is crucial. The bootstrap approach is likely better at preserving these empirical dependencies, as it resamples historical changes rather than modelling the curve parametrically.

In risk estimation, understanding how model performance varies across regimes is critical to avoid under- or overestimating risk during periods of stress. While all models experienced increased forecast errors and reduced correlations during volatile periods, the bootstrap model remained the most robust. The one-step DNS model deteriorated the most under stress, with wide prediction intervals, large MSE and low PCC.

### 8.2.2 Takeaways

The findings suggest that the bootstrap-based DNS method offers the most robust and reliable out-of-sample performance for yield curve forecasting. It performed well across all maturities and evaluation criteria, and remained comparably accurate even during periods of market stress. The method also produced narrower and more stable prediction intervals than the other models, while preserving key statistical features of the yield curves. This suggests that risk managers could obtain more precise estimates of future yields while accounting for extreme outcomes. Of course, the bootstrap's validity depends on the assumption that the future will be statistically similar to the past; if structural changes occur, it may need to be updated.

The two-step DNS performed reasonably well, except when reproducing the higher-order moments. The implementation is straightforward, so despite its limitations, it could still be helpful for routine forecasting and monitoring.

The one-step DNS model showed the weakest performance. It produced more volatile errors, wider prediction intervals, and struggled during turbulent periods. It also tended to overestimate volatility and exhibited negative autocorrelations, which were not observed in the empirical data. Negative autocorrelation implies frequent reversals in yield direction, which may cause the one-step DNS to underestimate the likelihood

of extended drawdowns, potentially leading to insufficient capital reserves. While the Kalman filter is theoretically appealing and can work well under the right conditions, it is sensitive to how the model is specified, and the model's forecasts can suffer if assumptions about the noise or initial values are incorrect. This indicates a gap between theoretical advantages and practical performance in real-world applications.

These findings also confirm that no model performs best across all evaluation criteria. Although the bootstrap method delivered the strongest overall performance, the two-step DNS also performed well, offering a practical balance between simplicity and accuracy. Risk analysts may benefit from using different models for different purposes. For instance, the two-step DNS could serve as a baseline model for routine forecasting, while the bootstrap may be better suited for exploring stress scenarios and tail risks.

Since all three methods share the same DNS structure, they are relatively easy to interpret. Each factor corresponds to a distinct yield curve component (level, slope, or curvature), making them suitable for stress testing. For example, adjusting the level factor allows analysts to examine the effect of parallel shifts on the entire curve. This interpretability improves the practical use of DNS-based models in regular forecasting and scenario analysis.

### 8.3 Limitations

Despite the overall successful application of DNS models, certain limitations of this study must be acknowledged. All three forecasting methods showed vulnerability to volatile markets. Significant, unanticipated events (such as central bank policy changes or geopolitical crises) led to forecast errors well beyond typical levels, showing that any purely data-driven model, parametric or not, can struggle with sudden structural changes.

Second, a structural constraint in the DNS framework used here was the fixed decay parameter  $\lambda$ . Although using a constant  $\lambda$  is common in DNS applications, it may not be optimal across different periods or market conditions. Yield curves that shift abruptly may be better captured with a time-varying  $\lambda$ , or an extended DNS model, such as the Nelson-Siegel-Svensson model, which includes an additional curvature term. It is also worth noting that the  $\lambda$  used here was proposed by F. Diebold and Li (2006), and perhaps results would have improved by another choice.

The Kalman filter implementation may also suffer from model misspecification or

suboptimal initialisation of state-space matrices. While the method is theoretically appealing, its practical performance depends heavily on this. Better implementation could improve its reliability, which was not explored in the current study.

Although robust in many respects, the bootstrap-based DNS method relies on the assumption that future yield changes will resemble those observed historically. In the case of sudden or prolonged market changes, the bootstrap's past samples may not adequately represent the new reality. This limitation was seen in the results, for example, during the 2022-2024 inflation surge.

## 8.4 Future Work

A clear next step would be to try a different choice of the decay parameter  $\lambda$ . Alternatives to the  $\lambda = 0.0609$  proposed by F. Diebold and Li (2006) include empirically calibrating  $\lambda$  based on historical estimates or using the average of fitted values across the sample, as in Huang (2021). Allowing  $\lambda$  to vary over time could also improve the model's ability to capture shifts in yield curve shape.

Another recommendation is to explore the use of full conformal prediction. This study implemented the split conformal method due to its simplicity and computational efficiency, but full conformal approaches may offer better performance.

Extending the model structure itself could also be interesting. The Nelson-Siegel-Svensson model, which adds a fourth factor, could capture additional curvature and improve fit, particularly during episodes of turbulent yield dynamics.

Given the promising performance of the bootstrap approach in preserving statistical properties, one could consider refining this method further. For instance, a conditional bootstrap — that also takes macroeconomic indicators or policy rate levels into account — could better capture the impact of exogenous information. Incorporating external factors may be particularly useful at the short end of the curve, which is highly sensitive to monetary policy and market sentiment.

# Appendix A

## Appendix

### A.1 State Space Models

The driving forces behind the evolution of economic variables are, for the most part, not directly observable or measurable. Standard VAR models cannot be used when the explanatory variables are not observable. Can use state space models instead.

State space models describes an observed time series  $\{y_t\}_{t=1}^T, y_t \in \mathbb{R}^{n_y}$  in terms of the underlying latent state vector  $\{x_t\}_{t=1}^T, x_t \in \mathbb{R}^{n_x}$ . The state variables evolve according to a stochastic process, usually a Markov process. The basic linear state space model consists of two equations:

#### The measurement equation

$$y_t = Hx_t + r_t, \quad r_t \sim \mathcal{N}(0, R)$$

describes how the observed data,  $y_t$  relate to the latent state variables,  $x_t$ , through the *measurement function*  $H \in \mathbb{R}^{n_y \times n_x}$ . The term  $r_t$  represent the measurement noise with covariance  $R \in \mathbb{R}^{n_y \times n_y}$ .

#### The transition equation

$$x_t = Ax_{t-1} + q_{t-1}, \quad q_{t-1} \sim \mathcal{N}(0, Q)$$

models the evolution of the state variables over time. The evolution of  $x_t$  is influenced by the previous state,  $x_{t-1}$  through the *state transition matrix*  $A \in \mathbb{R}^{n_x \times n_x}$ , as well as some process noise  $q_{t-1}$  with covariance  $Q \in \mathbb{R}^{n_x \times n_x}$  Kunst (2007).

To explicitly derive the recursion equations, the matrices  $H, A, R, Q$  need to be estimated. Likelihood-based *Kalman filtering* can be applied to do so.

## A.2 Kalman Filtering

Kalman filtering is a recursive method to estimate the latent variables of the linear state-space model by maximising the log-likelihood of the variables.

Assume that the observed time series  $Y_T = \{y_t\}_{t=1}^T, y_t \in \mathbb{R}^{n_y}$ , follows the state-space model A.1/A.1 and that we have an approximate set of model parameters:  $\delta = \{H, A, R, Q\}$ . Define the likelihood function for the given parameters  $\delta$  as  $f(Y_T; \delta)$ .

By Bayes' theorem, we can rewrite the likelihood as

$$\begin{aligned} f(Y_T; \delta) &= f(y_1 | \delta) f(y_2 | y_1, \delta) f(y_3 | y_2, y_1, \delta) \dots f(y_T | y_{T-1}, \dots, y_1, \delta) \\ &= \prod_{t=1}^T f(y_t | Y_{t-1}, \delta), \end{aligned}$$

where  $y_0 = \emptyset$  and  $Y_{t-1} = (y_1, \dots, y_{t-1})$  for  $t \geq 2$ . The log-likelihood function is then

$$\ln L(Y_T, \delta) = \sum_{t=1}^T \ln f(y_t | Y_{t-1}, \delta)$$

To construct the likelihood function we find the densities  $f(y_t | Y_{t-1}, \delta)$  for  $t = 1, 2, \dots, T$ . The Kalman filter can be used if the system is linear with Gaussian errors. Kunst 2007

The Kalman filter algorithm includes the following steps:

1. Initialisation
2. Prediction
3. Correction
4. Likelihood construction

Onwards,  $x_{t|s}$  will denote the estimate of  $x_t$  conditional upon the information available at a time  $s$ , and by  $P_{t|s}$  the associated error covariance matrix.

**Initialisation** To initialise the procedure, derive the best predictor of the initial state and an estimate of the covariance matrix. Specifically, define the initial state,  $x_{0|0}$ , along with its covariance matrix:

$$P_{0|0} = \mathbb{E}[(x_0 - x_{0|0})(x_0 - x_{0|0})^\top].$$

This step is straightforward if the process is stationary, as we can directly use the system's steady-state values Kunst 2007. In that case, we set  $x_{0|0} = x^*$  and  $P_{0|0} = P^*$ , such that

$$\begin{aligned} x^* &= Ax^*, \\ P^* &= AP^*A^\top + Q = [I - A \otimes A]^{-1} \text{vec}(Q). \end{aligned}$$

To initialise the filtering process, we set the previous state estimate to the initial values:

$$\begin{aligned} x_{t-1|t-1} &= x_{0|0}, \\ P_{t-1|t-1} &= P_{0|0}, \quad \text{with } t = 1. \end{aligned}$$

**Prediction** The prediction at time  $t$ , given the previous state estimate  $x_{t-1|t-1}$  and its covariance  $P_{t-1|t-1}$ , is computed as:

$$\begin{aligned} x_{t|t-1} &= Ax_{t-1|t-1}, \\ P_{t|t-1} &= AP_{t-1|t-1}A^\top + Q. \end{aligned}$$

The observation forecast is then given by  $y_{t|t-1} = Hx_{t|t-1}$ . We compute the forecast residual as:

$$\begin{aligned} u_t &= y_t - y_{t|t-1} = y_t - Hx_{t|t-1} \\ &= Hx_t + r_t - Hx_{t|t-1} \\ &= r_t - H(x_{t|t-1} - x_t), \end{aligned}$$

where we have used that  $y_t = Hx_t + r_t$ . Since  $r_t$  and  $x_t - x_{t|t-1}$  are independent, zero-mean Gaussian, it follows that  $u_t \sim \mathcal{N}(0, R + HP_{t|t-1}H^\top)$ .

To compute the likelihood, we need the conditional density  $f(y_t | Y_{t-1}; \delta)$  for  $t = 1, \dots, T$ . Since  $y_t = y_{t|t-1} + u_t$ , and  $u_t \sim \mathcal{N}(0, R + HP_{t|t-1}H^\top)$ , it follows that:

$$\begin{aligned} f(y_t | Y_{t-1}; \delta) &= f(u_t; \delta) \\ &= \frac{1}{\sqrt{(2\pi)^{n_y} |R + HP_{t|t-1}H^\top|}} \exp\left(-\frac{1}{2}u_t^\top(R + HP_{t|t-1}H^\top)^{-1}u_t\right). \end{aligned}$$

Hence, to evaluate  $f(y_{t+1} \mid y_t; \delta)$ , we must first update the state estimate and its covariance to obtain  $x_{t|t}$  and  $P_{t|t}$ , as described in the next step.

**Correction** Given an observation  $y_t$ , we update the predicted state  $x_{t|t-1}$  and its associated covariance  $P_{t|t-1}$  using the Kalman filter equations. For each time step  $t$ , compute the residual:

$$u_t = y_t - Hx_{t|t-1}.$$

The Kalman gain is computed using the covariance of  $u_t = HP_{t|t-1}H^\top + R$ :

$$K_t = P_{t|t-1}H^\top(H\Sigma_{t|t-1}H^\top + R)^{-1}.$$

Next, using  $K_t$  and  $u_t$ , we update the state and its covariance:

$$\begin{aligned} x_{t|t} &= x_{t|t-1} + K_t u_t, \\ P_{t|t} &= (I - K_t H)P_{t|t-1}. \end{aligned}$$

The residual ( $y_t - Hx_{t|t-1}$ ) captures the discrepancy between the expected and actual observation, and the Kalman gain adjusts how much of this discrepancy is used to correct the state estimate.

**Likelihood Construction** The densities computed in the prediction and update steps can now be used to calculate the likelihood

$$L(Y_T, \delta) = \prod_{t=1}^T f(y_t \mid Y_{t-1}, \delta).$$

**Likelihood Maximisation and Forecasting** Using the likelihood, maximum likelihood estimates of the parameters,  $\delta_{ML}$ , are computed numerically.

# References

- Nelson, Charles and Andrew F. Siegel (1987). “Parsimonious Modeling of Yield Curves”. In: *Journal of Business* 60.4, pp. 473–489.
- Vovk, Vladimir, Alex Gammerman, and Glenn Shafer (2005). *Algorithmic Learning in a Random World*. Springer Science & Business Media.
- Diebold, Francis and Canlin Li (2006). “Forecasting the term structure of government bond yields”. In: *Journal of Econometrics* 130.2, pp. 337–364. URL: <https://www.sciencedirect.com/science/article/pii/S0304407605000795>.
- Diebold, Francis, Glenn Rudebusch, and Boragan Aruoba (2006). “The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach”. In: *Journal of Econometrics* 131.1-2, pp. 309–338. DOI: 10.1016/j.jeconom.2005.01.011. URL: <https://www.sciencedirect.com/science/article/pii/S030440760500014X>.
- Kunst, Robert (2007). “State Space Models and the Kalman Filter”. In: URL: <https://homepage.univie.ac.at/robert.kunst/statespace.pdf>.
- Christensen, Jens, Francis X. Diebold, and Glenn D. Rudebusch (2009). “An arbitrage-free generalized Nelson–Siegel term structure model”. In: *The Econometrics Journal* 12.3, pp. C33–C64. URL: <https://www.sas.upenn.edu/~fdiebold/papers/paper80/CDR2.pdf>.
- Bohner, Martin (2011). *Fixed Income Models: Winter 2010/2011 Lecture Notes*. Lecture notes. Version from January 31, 2011. Rolla, Missouri 65409-0020, USA. URL: <http://web.mst.edu/~bohne>.
- Christensen, Jens, Francis Diebold, and Glenn D. Rudebusch (2011). “The affine arbitrage-free class of Nelson–Siegel term structure models”. In: *Journal of Econometrics* 164.1, pp. 4–20. URL: <https://www.sciencedirect.com/science/article/pii/S0304407611000388>.
- Angelopoulos, Anastasios and Stephen Bates (2021). *A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification*. arXiv: 2107.07511 [stat.ML]. URL: <https://arxiv.org/abs/2107.07511>.

- Huang, Zhe (2021). “Fitting Yield Curve with Dynamic Nelson-Siegel Models: Evidence from Sweden”. Master’s Thesis. Uppsala, Sweden: Uppsala University. URL: <https://uu.diva-portal.org/smash/get/diva2:1593645/FULLTEXT01.pdf>.
- Barber, Rina Foygel et al. (2023). “Conformal Prediction Beyond Exchangeability”. In: *The Annals of Statistics* 51.2, pp. 816–845. DOI: 10.1214/23-AOS2276. URL: <https://projecteuclid.org/journals/annals-of-statistics/volume-51/issue-2/Conformal-prediction-beyond-exchangeability/10.1214/23-AOS2276.full>.