Forecasting the Dynamics of the Term Structure of Interest Rates under Different Economic Scenarios: a Dynamic Latent Factor Approach

by
J.G. Kremer (2032227)

Supervisor:
Prof. Dr. B. Melenberg

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Quantitative Finance and Actuarial Science

Tilburg School of Economics and Management
Tilburg University

January 20, 2024

Abstract

This study examines and evaluates the forecasting performance of two extensions of the dynamic Nelson-Siegel model when applied to Treasury yields in the United States and 6-month Euribor swap rates in the Euro Area. These extensions involve integrating a measure for the central bank's balance sheet alongside other fundamental macroeconomic variables. Additionally, the study explores the model's sensitivity to forecasting performance when addressing non-stationarity in underlying processes. The analysis encompasses various forecasting horizons, revealing consistent model performance in both regions. The extension incorporating a measure for the central bank's balance sheet demonstrates improved forecasting accuracy across the entire term structure for longer forecast horizons. For the 1-month forecast horizon, the model corrected for non-stationarity without macro variables exhibits the best forecasting performance. Beyond forecasting, the study leverages the most effective model for scenario analysis, providing valuable insights into potential interest rate movements under different economic scenarios.

Contents

1 Introduction

The current state of the global economy is characterized by a high degree of uncertainty and volatility, accompanied by elevated levels of inflation, rising market interest rates, and heightened geopolitical tensions. These prevailing economic conditions have presented significant challenges in accurately predicting the future trajectory of the economy, particularly with respect to the term structure of interest rates. In response to these economic conditions, the European Central Bank (ECB) has implemented a series of successive interest rate hikes, resulting in the main refinancing operations (MRO) rate reaching an unprecedented record high of 4.00%. Similarly, the Federal Reserve (Fed) in the United States has raised its funds rate to 5.5%, marking the highest level in over 22 years. This stands in stark contrast to the zero interest rates observed during the height of the pandemic. Both the US and the Euro Area (EA) have experienced historically rapid increases in interest rates. In addition, both regions have observed persistent inverted interest rate curves, which is a strong indicator of an upcoming recession and reflects investors' negative expectations about the economy's outlook. In addition to the interest rate hikes, central banks in both the US and the EA have started a process of reducing their balance sheets through quantitative tightening (QT), aiming to further slow down the economy. The current monetary tightening cycle represents a significant shift from the quantitative easing (QE) policies of the pandemic era and the global financial crisis. The increasingly complex economic landscape and the growing reliance on unconventional monetary policy have heightened the unpredictability of interest rates, posing challenges for homeowners, investors, risk managers, and pension funds, who rely heavily on accurate interest rate forecasts.

Over the last four decades, the field of interest rate modeling has experienced significant progress in terms of both the development of theoretical models and the econometric techniques used to estimate these models. In contrast to one accepted standard, such as the Black and Scholes (1972) model for equity option pricing theory, the field of interest rate modeling is characterized by three prominent frameworks: the equilibrium models, the arbitrage-free models, and the empirical models. The differences in these frameworks reflect the different motives of the researchers who build these models.

The first category of models, which typically are affine models, is grounded on economic theory, with a primary focus on the dynamics of the economy. These models are stochastic interest rate models that model the term structure as a function of a number of economic factors, thereby providing a theoretical depiction of the term structure. Prominent contributions are the one-factor models from Vasicek (1977) and Cox et al. (1985), where the entire term structure is constructed from modeling the dynamics of the short rate. Heston (1993) and Brennan and Schwartz (1982) are some of the notable extensions in this category of models. The work of Duffie and Kan (1996) can be considered the most general model within this category, encompassing the aforementioned models as special cases.

In contrast to equilibrium models, the arbitrage-free class of models does not incorporate economic dynamics explicitly but assumes that the term structure follows a random process. These models are calibrated using real data and focus on optimally fitting the observed term structure at a point in time, thereby ensuring the absence of arbitrage opportunities. This makes the no-arbitrage category more suited for derivative pricing. The fundamental assumption underlying these models is that the market term structure is arbitrage-free. Prominent contributions to this category of models include Ho and Lee (1986), Heath et al. (1992), and Hull and White (1990), where the latter

extend the model of Vasicek by introducing time-dependent parameters.

Both the equilibrium and arbitrage-free classes of models are more concerned with in-sample term structure fitting. Their out-of-sample forecasting capabilities are rather limited, as demonstrated by Duffee (2002). This led to a shift in focus towards the third class of models. The empirical class of models embraces a data-driven approach and has an explicit focus on out-of-sample forecasting. This emphasis on empirical validation and forecasting performance has led to the development of models like the class of models of Nelson and Siegel (1987), which effectively capture patterns and regularities observed in historical interest rate data. The promising forecasting performance of the Nelson and Siegel model led to many extensions in this direction, e.g., Svensson (1994), and Björk and Christensen (1999) extended the model by incorporating an extra factor which led to enhanced performance, both in-sample as out-of-sample according to Pooter (2007). Diebold et al. (2006) transformed the model into a state space framework and characterized the dynamic interactions of the macroeconomy and the term structure by augmenting the model with macroeconomic variables. Christensen et al. (2009) introduced an arbitrage-free Nelson-Siegel model, connecting theoretical foundations and practical applications.

In both the US and the EA, extensive research has been conducted to examine the impact of large-scale asset purchases by the respective central banks on the term structure of interest rates. Consistently, various studies have revealed significant outcomes, indicating that the announcement and implementation of QE policies notably affect the back end of the term structure. Conclusions drawn from these studies uniformly emphasize that the effects on the front end of the term structure are either insignificant or minimal, resulting in a flattening of the term structure. For instance, Eser et al. (2019) identified that the APP persistent reduced the 10-year rate by 95 basis points (bsp). Gambetti and Musso (2020) and van Dijk, Dubovik, et al. (2018) replicated similar outcomes for interest rates in the EA. Dröes et al. (2017) even found evidence suggesting that the substantial reduction in interest rates due to QE led to a significant deviation of bond prices from their underlying fundamental value. In the context of the US, the first two rounds of QE by the Fed resulted in a decline of 107 basis points bsp in the 10-year Treasury yield and 25 basis points in the 1-year Treasury yield (Krishnamurthy & Vissing-Jorgensen, 2011), indicating a flattening of the yield curve. Also, Gagnon, Raskin, Remache, and Sack (2011) identified a similar effect, causing the Treasury yield curve to flatten. These effects align with the rationale that QE policies primarily target assets with prolonged maturity.

This study is centered on the modeling and forecasting of US Treasury yields and EA swap rates by incorporating macroeconomic information into the modeling framework. Given the recognized challenges in accurate point forecasting and the inherent volatility of economic dynamics, this study places a significant emphasis on generating the term structure under different economic scenarios. By integrating macroeconomic variables into the modeling framework, with a particular focus on monetary policy decisions, this study establishes the relationship between the term structure and the dynamics of the macro variables. This relationship allows for the generation of future scenarios by adjusting the outlook for these variables. To enhance the relevance of the analysis, recent data up to November 2023, spanning maturities from 3 months to 30 years, is included

To achieve the study objectives, the initial step involves conducting explanatory data analysis on the interest rate data for both regions. This assessment aims to identify potential differences in the dynamics and characteristics of the data. Descriptive statistics and autocorrelation at various displacements for each maturity are examined. Additionally, the study evaluates the descriptive

statistics of the empirical level, slope, and curvature for each region. Following this, a similar analysis is performed for selected macro variables. The criteria for choosing these variables include maintaining parsimony in the model by selecting a minimal set of fundamental variables that capture the primary dynamics in the macroeconomy. Factors such as interpretability are crucial as the study emphasizes scenario analysis. The chosen macro variables are those for which there is a general consensus about their outlook, ensuring the ability to derive a reliable and plausible main scenario. This criterion leads to the inclusion of macro variables with projections issued by the central bank.

Next, following the methodology of Diebold et al. (2006), this study explores various specifications of the dynamic Nelson-Siegel (DNS) model in state-space form. The model specifications considered include (i) Yield-Only (YO), (ii) Yield-Macro 3 (YM3), and (iii) Yield-Macro 4 (YM4), which represent models with no macro variables, three macro variables, and four macro variables, respectively, where the fourth macro variable measures the central bank's balance sheet. To demonstrate the efficacy of the state-space formulation, the approach of Diebold and Li (2006) is incorporated for models without macro variables. Each model's in-sample and out-of-sample performance is subsequently evaluated using different error metrics. In both regions, there is suggestive evidence that the YM4-KF provides the best in-sample fit, particularly excelling for mid-term maturities. The estimated factors exhibit a strong correlation with their empirical counterparts, reinforcing the interpretation of the factors as representing the level, slope, and curvature. The out-of-sample forecast performance of models incorporating macro variables is superior to those without macro variables for longer forecast horizons based on mean Root Mean Squared Error (RMSE), with the YM4-KF model demonstrating superiority. However, no models outperform the random walk for the 1-month forecast horizon.

To further enhance forecasting accuracy, an assessment is conducted to verify the adherence to model assumptions. The implicit assumption of the state space methodology is that the state vector follows a stationary process (Durbin & Koopman, 2012). The stationarity of each time series is evaluated using a combination of the Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test, and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, presented by Dickey and Fuller (1981), Kwiatkowski et al. (1992), and Phillips and Perron (1988). Non-stationary processes are then transformed into stationary processes through the application of a first-order difference transformation. Subsequently, the forecasting performance of each model is evaluated using the same procedure as before. Strong evidence suggests that variables exhibit non-stationarity. After applying the first-order difference transformation, there is evidence that the processes become stationary. In both regions, the models outperform their equivalent models estimated on level data for the 1-month horizon based on mean RMSE. However, for longer forecast horizons, the models estimated on level data demonstrate superiority.

Scenarios are based on the YM4-KF model. Horizons up to 5 years are considered. The main scenario is based on the projections of the respective central bank. Additionally, a recession and an economic expansion scenario are considered. The evolution of the term structure under each scenario aligns with the economic expectations, with the entire term structure of the expansion and recession scenario positioned both above and below that of the main scenario, respectively. These two alternative scenarios can be interpreted as an upper and lower bound since they represent more extreme situations. From the main scenario, it can be concluded that the inversion of the term structure stabilizes and transitions to a standard upward-sloping curve in mid-2024 for both regions.

This study's contribution to the existing literature on term structure modeling is threefold, with each contribution being distinct yet interrelated. Firstly, a metric for the central bank's balance sheet is incorporated into the modeling framework for both the US and the EA. This integration aims to enhance the forecasting accuracy of interest rate movements and characterize the interplay between the balance sheet of the central banks and the term structure of interest rates. Secondly, this study examines the stationarity conditions within the term structure models, implementing appropriate adjustments to ensure robustness in the model's predictive capability. By addressing the issue of stationarity, this study tries to provide insights into the stability of the model and enhance forecast performance. The last significant contribution arises from the ability to conduct scenario analysis, providing insights into the various potential paths of the interest rate term structure under different economic scenarios. This stands in contrast to relying solely on precise point forecasting.

The remainder of this paper is structured as follows: Section 2 describes the main concepts relevant to the study, while Section 3 explores the most prominent extensions of the class of Nelson-Siegel models. Section 4 outlines the methodology, and Section 5 covers explanatory data analysis. Section 6 introduces a measure for the balance sheet, followed by discussions on in-sample/out-of-sample performance in Section 7. Section 8 addresses stationarity conditions and compares forecasting results. Section 9 presents term structure trajectories under diverse economic scenarios. A concluding summary is provided in Section 10, and Section 11 outlines limitations and proposes directions for future research

2 Concepts and background

This section provides a comprehensive overview of the four main different term structure representations and an in-depth exploration of the conventional and unconventional monetary policies maintained by the central banks to ensure economic and financial stability.

2.1 Term structure descriptions

The term structure of interest rates, or term structure, plays a fundamental role in understanding the dynamics of financial markets and has a guiding role in investment decisions and the assessment of the current state of the economy. The term structure describes the cross-sectional relationship between bonds with varying time to maturities and their corresponding interest rates at a specific moment in time. The time to maturity is defined as the time remaining until a financial contract expires and is denoted as $\tau = T - t$, where T represents the date at which the financial contract is due, and t represents the date at which the contract is initiated. By assessing the term structure, market participants gain valuable insights into market expectations, risk assessments, and pricing of fixed-income securities. There are multiple ways to represent the term structure of interest rates. The four most common characterizations are the discount, yield, forward, and swap curves. These four representations, while closely related, offer distinct perspectives. Each curve has representations based on discrete and continuous compounding. For theoretical purposes, continuous compounding is used throughout this study. The primary focus of this research is on constructing zero-coupon yield curves for the US and 6-month Euribor swap curves for the EA.

The simplest instrument for determining the interest rate corresponding to a specific time to maturity τ, is a default-free zero-coupon bond. This bond guarantees a fixed payment for time to maturity τ and does not generate any other cash flows. In practice, most bonds typically have coupon payments. Such bonds can be seen as portfolios of zero-coupon bonds.

The discount factor, denoted as $P_t(\tau)$, can be understood as the price of a zero-coupon bond at time t that pays 1 at maturity τ and is used to determine the present value of a series of future cash flow. The discount curve is derived from the values of the discount factors at different maturities τ. This representation is considered the most fundamental representation of the term structure of interest rates (Schumacher, 2020). The yield, denoted as $y_t(\tau)$, is the implied zero-coupon rate. The most prevalent way to depict the term structure is by plotting $y_t(\tau)$ against τ, resulting in the yield curve. The relationship between the discount factor and the corresponding continuously compounded interest rate is given by

$$P_t(\tau) = e^{-y_t(\tau)\tau} \iff y_t(\tau) = -\frac{1}{\tau}\ln(P_t(\tau)). \tag{2.1}$$

The short-term interest rate or the instantaneous rate, denoted by r_t, is found by taking the limit as τ approaches zero. This is the yield on a loan one would gain if it would invest its money for an infinitesimally short period of time. The three-month or six-month rate is often used as a proxy for the short rate rather than the overnight rate since the three-month and six-month rates are less volatile and more liquid, making them a more stable proxy rate.

The forward rate, denoted as $f_t(T_1, T_2)$, refers to the implied interest rate at time t that a bond will yield if an investor purchases a bond at a specific time T_1 in the future and holds the bond

until maturity T_2. The relationship between the forward rate and the yields is given by

$$f_t(T_1, T_2) = \frac{y_t(T_2)(T_2 - t) - y_t(T_1)(T_1 - t)}{T_2 - T_1}. \tag{2.2}$$

From the forward rate $F_t(T_1, T_2)$, the instantaneous forward rate $f_t(\tau)$ can be determined by setting T_1 equal to T and T_2 equal to $T + \Delta T$, and then taking the limit as ΔT approaches zero. This process establishes a direct relationship between the instantaneous forward rate, discount rate, and yield. These relationships are expressed as:

$$f_t(\tau) = \frac{-P'(\tau)}{P(\tau)} \iff P_t(\tau) = e^{-\int_0^\tau f_t(u)du} \tag{2.3}$$

$$f_t(\tau) = (\tau y_t(\tau))' \iff y_t(\tau) = -\frac{1}{\tau}\int_0^\tau f_t(u)du \tag{2.4}$$

The forward rate acts as an arbitrage mechanism by making investors indifferent between entering a rollover strategy and purchasing a long-term bond. From equation 2.4, one can state that the yields can be interpreted as the cumulative average of the forward rates.

The swap curve is the last representation of the term structure that will be discussed. The par swap rate, denoted by $s_t(\tau)$, is the rate at which the present value of the swap is equal to zero at initiation. The par swap rate can be expressed in each of the aforementioned rates. However, expressing swap rates directly in terms of yields is not straightforward and lacks a simple intuitive formula. The mapping of swap rates across different the maturities τ results in the swap curve t.

$$s_t(\tau) = \frac{1 - P(\tau)}{\int_0^\tau P(u)du} \tag{2.5}$$

$$s_t(\tau) = \frac{\int_0^\tau P(u)f(u)du}{\int_0^\tau P(u)du} \tag{2.6}$$

2.2 Balance Sheet Policy

In response to evolving economic conditions, central banks employ various tools to effectively steer the economy. The policy rate is the primary tool employed by central banks to regulate the supply of money and influence the inflation rate in the economy. The policy rate is an uncollateralized overnight rate that acts as a benchmark for all other interest rates in the financial markets. By adjusting this policy rate, central banks want to lead the short-term end of the term structure, letting longer-term rates be determined mainly by market forces (Braun, 2017). When faced with inflationary pressures, central banks typically raise the policy rate. They aim to slow down the economy by increasing interest rates and reducing the available credit and liquidity in the market. Conversely, central banks lower the policy rate during economic downturns to stimulate economic activity by reducing rates. However, in recent decades, the effectiveness of traditional monetary policy has been challenged by persistently low nominal interest rates. As interest rates approach their effective lower bound, central banks face limited room for further interest rate reductions without adversely affecting credit supply, economic activity, and inflation. This constraint led to the introduction of a new monetary policy tool called balance sheet policy. Balance sheet policy encompasses a range of strategies that influence the central bank's balance sheet to achieve specific monetary policy objectives. One prominent form of balance sheet policy is large-scale asset purchase programs, commonly known as Quantitative Easing (QE). QE involves the central bank acquiring significant amounts of assets, such as government bonds, asset-backed securities, and equities, from financial institutions. This injection of liquidity into the market aims to lower yields,

mainly on the long end of the curve, and expand the money supply, which in turn should stimulate economic activity (Krishnamurthy and Vissing-Jorgensen, 2011). Quantitative Tightening (QT) is the opposite of QE, and aims to increase interest rates by reducing the central bank's balance sheet. This is done by offloading assets acquired during QE or by not reinvesting maturing securities.

The Bank of Japan (BoJ) pioneered in employing this new form of monetary policy. In 1999, the BoJ introduced negative interest rates to combat domestic deflation. To further stimulate the economy, the BoJ announced the first QE program, which lasted until 2006. Many central banks followed, with the first big adaption of QE in the aftermath of the global financial crisis. In 2009, the US announced its first large-scale asset purchase program (LSAP1) in response to the financial crisis. This first round of QE expanded the balance sheet of the federal reserve from 877$ billion to 2.3$ trillion due to the purchases of mortgage-backed securities, bank debt, and Treasury notes over the course of six months. Mainly, long-term interest-rate products were acquired, which led to an increase in their prices and a decrease in the Treasury yields due to the inverse relationship. In total, three more QE programs followed, with the last one concluding in 2020 as a response to the COVID-19 pandemic.

2.3 ECB's conventional monetary policy measures

The ECB is the central bank for the Eurozone or EA countries. The Eurozone consists of European Union (EU) member countries that have adopted the euro as their official currency. The ECB is responsible for monetary policy within the EA, and its primary objective is to maintain price stability, ensuring that inflation remains low and stable. The EA countries that have adopted the euro share a common monetary policy set by the ECB, while fiscal policies remain the responsibility of each individual member country.

The euro short-term rate (ESTR) is the uncollateralized overnight rate, serving as the risk-free rate benchmark for the Eurozone. It replaced the Euro Overnight Index Average (EONIA) rate in 2019 and is derived from actual transactions from overnight borrowing rates in the interbank and wholesale markets. By effectively managing this market's short-term interest rate, the ECB transmits its monetary policy measures to the economy. The ECB employs the corridor system to steer ESTR towards its desired level. The corridor exists out of the three key interest rates: the main refinancing operations (MRO) rate, the deposit facility (DF) rate, and the marginal lending facility (MLR) rate. The MLR and the DF rate are the rates for which a bank can borrow or store liquidity overnight, respectively. The MLR serves as the upper limit of the corridor and the DF rate serves as the lower limit of the corridor, as shown in figure 1. The MRO rate, often referred to as the policy rate of the ECB, operates in between the corridor, and it represents the rate at which banks can lend money from the central bank in the medium term.

2.4 ECB's unconventional monetary policy measures

In response to the global financial crisis, the European Central Bank (ECB) initiated two covered bond purchase programs (CBPP1 and CBPP2) to address financial instability and promote economic growth (Smith, 2020). These programs aimed to enhance liquidity in the covered bond market, alleviate funding challenges faced by banks, and mitigate poor credit conditions. The ECB purchased €76.4 billion worth of assets under these programs, which closely resembled QE measures. However, the ECB's then-current president explicitly stated that the bank did not engage

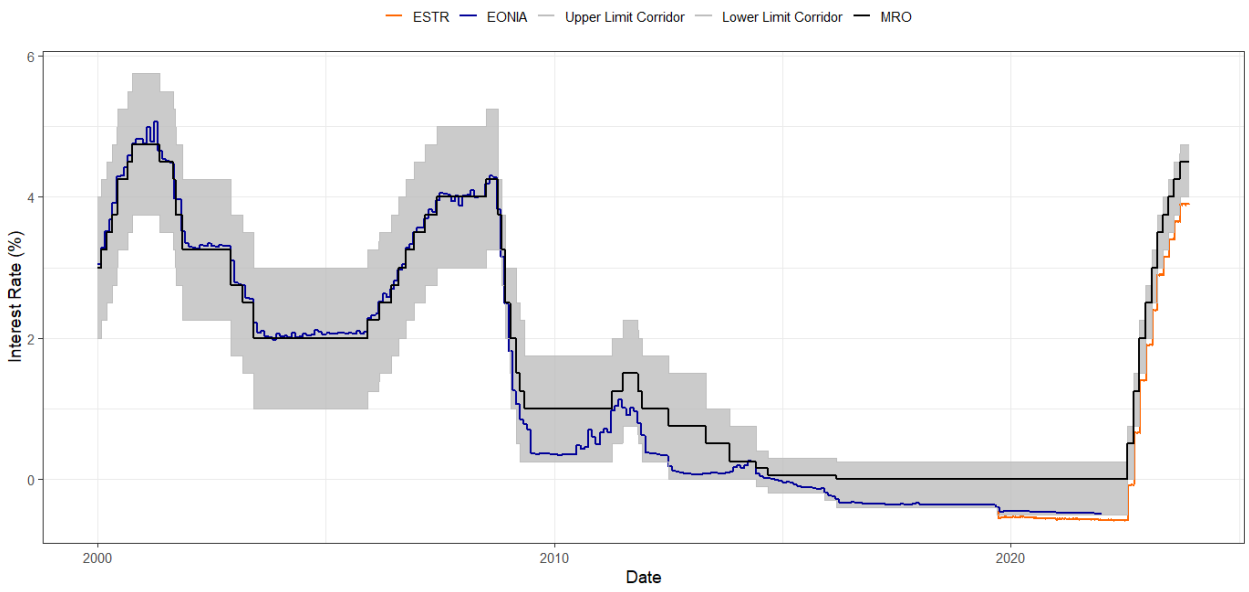


Figure 1: Corridor system for the three policy rate of the ECB

in QE and instead labeled these actions as "enhanced credit support" (Trichet and Papademos, 2009). In 2010, the escalating sovereign debt crisis in the euro area triggered the ECB to establish the Securities Market Program (SMP). The primary objective of the SMP was to prevent contagion of financial instability and provide liquidity by purchasing sovereign bonds of distressed countries. While this program shared similarities with QE, it was not classified as such due to the ECB's active sterilization efforts, which offset the liquidity injection of the SMP and prevented an overall expansion of the money supply. The aforementioned programs have been terminated and were the first encounters with unconventional monetary measures for the ECB. As the policy rate approached the zero lower bound in 2014, the ECB once again resorted to unconventional monetary measures to stimulate the economy. The bank launched its asset purchase program (APP), a comprehensive program encompassing four separate packages with distinct starting dates, terms, and asset classes. The ECB systematically purchased assets, beginning with asset-backed securities (ABSPP) and covered bonds (CBPP3). In 2015, the ECB initiated the public sector purchase program (PSPP), which involved buying government bonds, followed by the purchasing of corporate bonds under the corporate sector purchase program (CSPP). Under the APP the ECB bought for EUR 2580 billion worth of assets over the course of 3.5 years. In December 2018 the APP paused for a year. During this pause principal payment of maturity securities were reinvested but no new net purchases were done. In October 2019 the APP was relaunched during the outbreak of Covid-19 crisis. In addition to the relaunch of the APP, the ECB introduced the Pandemic Emergency Purchase Program (PEPP). This temporary asset purchase progam injected EUR 1850 billion in the system. Due to the sharp increase in inflation above the 2% in 2022, the ECB changed its monetary policy from QE to a soft from of QT, through slowly offloading assets acquired during these programs. This is done by limiting reinvestments of maturing securities.

In addition to the APP and PEPP, the ECB also launched multiple Targeted Longer-Term Refinancing Operations (TLTROs) starting of 2014. These operations, categorized as Credit Easing, provide long-term loans to euro area banks that meet specific criteria at a discounted interest rate. The goal is to stimulate bank lending to the economy by offering incentives to banks in the form of

favourable borrowing conditions.

3 Class of Nelson-Siegel models

In this section, the relevant technical details of the classical Nelson-Siegel model and its significant extensions are discussed, together with some strengths and weaknesses of each of the models.

3.1 Original Nelson-Siegel model

The Nelson-Siegel class of models is a family of statistical models for fitting the term structure of interest rates. The original model, proposed by Nelson and Siegel (1987), presents a straightforward and adaptable parsimonious framework designed to fit the cross-section of interest rates at any given time. This model demonstrates the capability to capture the diverse shapes commonly observed in yield curves, including monotonic, humped, and S-shaped. This model decomposes the term structure into several components: the three latent factors $\beta_1, \beta_2, \beta_3$, and the parameter λ. Each of the three latent factors is weighted by their factor loadings $1, (\frac{1-e^{\tau\lambda}}{\tau\lambda}), (\frac{1-e^{\tau\lambda}}{\tau\lambda} - e^{-\lambda\tau})$ respectively, where τ represent the time to maturity. Equation 3.1 depicts this relationship between the yields and the time to maturity as proposed by Nelson and Siegel (1987).

$$\begin{aligned} y(\tau) &= \beta_1 + \beta_2 \cdot (\frac{1-e^{-\tau\lambda}}{\tau\lambda}) + \beta_3 \cdot (\frac{1-e^{-\tau\lambda}}{\tau\lambda} - e^{-\lambda\tau}) \\ &= \beta_1 + (\beta_2 + \beta_3) \cdot (\frac{1-e^{-\tau\lambda}}{\tau\lambda}) - \beta_3 \cdot e^{-\lambda\tau} \end{aligned} \tag{3.1}$$

Each component describes a different aspect of the behavior of the yield curve. The parameter λ controls the exponential rate of decay, and it determines for which maturity the loading on β_3 obtains its maximum value as illustrated in figure 2. Large values of λ result in a fast decay and shorten the maturity at which the loading on β_3 obtains its maximum. Conversely, smaller values of λ result in a slower decay of the loadings on the factors.

Based on the behavior of the loadings, each of the three factors can be described as the long-term term factor, short-term factor, and middle-term factor, respectively. β_1 is considered the long-term factor since the loading on β_1 equals 1 and does not decay in the limit. This makes β_1 the dominant factor in determining the shape of the yield curve at the long end, as the loadings on the other factors converge to zero. This can be shown by letting the maturity go to ∞. β_2 can be described as the short-term factor since the loading on β_2 starts at 1 and decays relative quickly to zero as the maturity increases. This makes the parameter β_2 mainly dominant in determining the shape of the yield curve for short maturities. The loading on β_3 exhibits a bell-shaped curve, with their peak in the middle and gradually going to zero for short and long maturities. This pattern suggests that the function is primarily active for maturities in the middle range.

3.2 Dynamic Nelson-Siegel

Diebold and Li (2006) extended the previous work of Nelson and Siegel (1987) by introducing a dynamic framework to capture the evolution of interest rates and focusing specifically on forecasting the yield curve. They maintained a two-step procedure. Initially, they derived a series for each of the

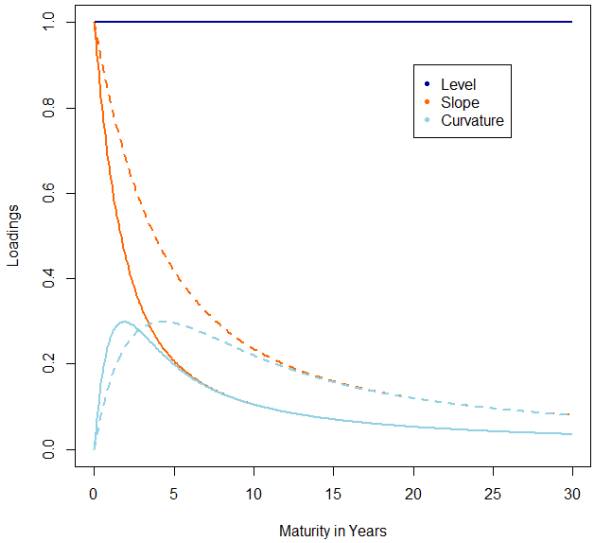


Figure 2: Behaviour of the loadings of the factors in the Nelson-Siegel model under different λ. Dotted line represents $\lambda = 0.035$. Solid line represents $\lambda = 0.08$

three latent factors corresponding to distinct time points t by applying the Nelson Siegel to cross-sectional interest rate data. Subsequently, in the second step, the dynamics of these individual factors were modeled using a first-order autoregressive model. Diebold and Li (2006) recognized that these three latent factors, initially classified as β_1, β_2, and β_3, effectively represent the yield curve's level, slope, and curvature, respectively. This conceptualization led to a reformulation of the original model, given by:

$$y_t(\tau) = L_t + S_t \cdot \left(\frac{1 - e^{-\tau\lambda}}{\tau\lambda}\right) + C_t \cdot \left(\frac{1 - e^{-\tau\lambda}}{\tau\lambda} - e^{-\lambda\tau}\right) \tag{3.2}$$

Here $y_t(\tau)$ represents the yield at time t for maturity τ, while L_t, S_t, and C_t correspond to the level, slope, and curvature factors, respectively. Here is the level factor identified as the long-term factor. This can be verified by letting τ approach to infinity (as depicted in equation 3.3). It represents the baseline or average level of interest rates in the yield curve and is empirically defined as $y_t(\tau_N)$, where τ_N represents the longest maturity in months in the term structure one considers. The slope factor, defined empirically as the difference between the interest rates at τ_N months (long-term) and τ_1 months (short-term), measures the steepness or flatness of the curve. The theoretical definition of slope, defined as the difference between the interest rates at infinite maturity and 0 maturity, precisely corresponds to $-S_t$. In this model, the theoretical short-term rate r_t is equivalent to the sum of the level and slope factors, equation 3.4. Furthermore, Diebold and Li defined the curvature of the yield curve as 2 times the interest rate at 24 months minus the sum of the interest rates at 3 months and τ_N months. They demonstrated that this curvature measure is closely linked to the medium-term factor, denoted as C_t. Figure 3 visualizes the decomposition of an upward-sloping interest rate curve into the three latent factors.

$$\lim_{\tau \to \infty} y_t(\tau) = L_t \tag{3.3}$$

$$r_t = \lim_{\tau \to 0} y_t(\tau) = L_t + S_t \tag{3.4}$$

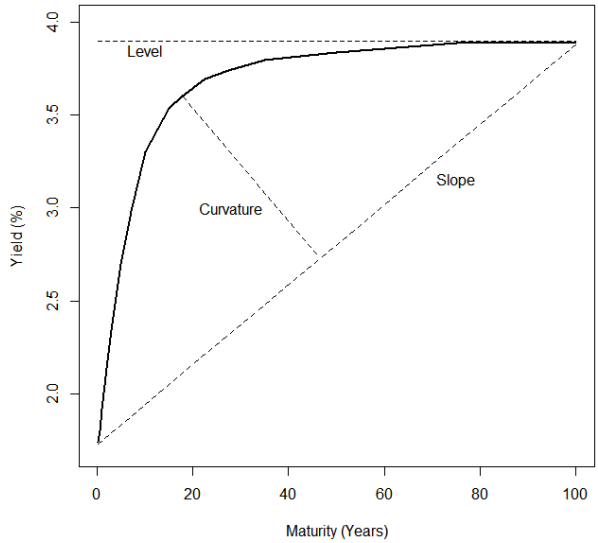


Figure 3: Term structure decomposition into level, slope, and curvature

3.3 Extended Dynamic Nelson-Siegel in State Space Form

Shortly after Diebold and Li published their model, Diebold et al. introduced a revised version incorporating a state-space framework to capture the dynamic relationship between latent factors and the yield curve. Parameters within this framework are estimated using the Maximum Likelihood Estimation (MLE) methodology, with state estimates provided by the Kalman Filter. A more comprehensive examination of the state-space framework and the estimation procedure will be provided in Section 4.2. This revised approach improves upon the two-step approach by enabling the simultaneous fitting of the yield curve at any time $t$along with estimating the latent factors and their dynamics. In addition, this one-step approach yields accurate inferences via established statistical theory.

The vector-matrix notation of the dynamic Nelson-Siegel model in state-space form is given below. Equation 3.5 represents the measurement or observation equation, and equation 3.6 represents the state or transition equation.

$$y_t = \Lambda f_t + \epsilon_t. \tag{3.5}$$

$$(f_t - \mu) = \Phi(f_{t-1} - \mu) + \eta_t, \tag{3.6}$$

In the specified framework, y_t is defined as an m dimensional vector of yields for m different maturities and f_t is an n dimensional vector corresponding to n factors. Λ_t an $m \times n$ matrix, denoting the loadings on the factors, ϵ_t is an m dimensional vector capturing the measurement error in the measurement equation. In the transition equation, A is a $n \times n$ matrix corresponding to the parameter matrix of the VAR model, and μ is a n dimensional vector representing the constant mean of the factors. Lastly, η_t is an n dimensional vector capturing the white noise transition errors. It is worth noting that the stationarity of the factors is implicitly assumed since a constant mean is employed

In addition, Diebold and Li (2006) extended the DNS model by incorporating macroeconomic factors to characterize the dynamic interaction between the macroeconomy and the yield curve. A minimal number of factors were introduced to maintain model parsimony while still capturing the essential dynamics in the macroeconomy. This version of the model is referred to as the Yield-Macro-3 (YM3) model. The added macroeconomic factors include manufacturing capacity utilization (CU_t), the federal funds rate (FFR_t), and annual price inflation ($INFL_t$). Their study uncovered empirical support for a bidirectional relationship between macroeconomic variables and the yield curve. Notably, the influence of macroeconomic factors on yield curve dynamics appeared to be more pronounced. This finding underscores the significant impact of economic indicators on interest rate movements and highlights the interconnected nature of macroeconomic conditions and bond markets. This model is discussed in full detail in Section 4.1

3.4 Svensson, Arbitrage-Free Nelson Siegel and Time Varying Parameters

This subsection introduces three additional extensions; however, they are not directly employed in this study and are therefore discussed in less detail. Each of these extensions can be formulated in state-space form.

3.4.1 Nelson-Siegel-Svensson

Svensson (1994) extended the original Nelson-Siegel model by introducing a fourth factor β_4, with an additional decay parameter λ_2, resulting in:

$$y(\tau) = \beta_1 + \beta_2 \cdot (\frac{1 - e^{-\tau\lambda_1}}{\tau\lambda_1}) + \beta_3 \cdot (\frac{1 - e^{-\tau\lambda_1}}{\tau\lambda_1} - e^{-\lambda_1\tau}) + \beta_4 \cdot (\frac{1 - e^{-\tau\lambda_2}}{\tau\lambda_2} - e^{-\lambda_2\tau})$$

The incorporation of an additional factor and a decay parameter enhances the flexibility and fitting capability of the model. The loading on β_4 mirrors that of β_3, introducing a second medium-term component that effectively captures the curvature of the yield curve, leading to a more accurate fit for yield curves with multiple local extrema along the maturity spectrum. The limiting behavior of yields remains unchanged due to the structural similarity of the loadings on β_3 and β_4. However, the addition of the extra decay parameter makes this extension of the model more complex and highly non-linear, making estimation of the model extra demanding, see Bolder and Stréliski (1999). In addition, the multicollinearity condition is violated when λ_2 takes on values that are close to λ_1, making estimation and intractability of the factors inefficient and unclear.

3.4.2 Arbitrage-Free Nelson-Siegel

Christensen et al. (2009) overcomes the biggest drawback of the Nelson-Siegel model by introducing an extension to the DNS model, which makes it arbitrage-free. The Arbitrage-Free Nelson-Siegel (AFNS) model still belongs to the affine class of term structure models and incorporates a yield-adjustment term $-\frac{C(\tau)}{\tau}$, resulting in

$$y(\tau) = \beta_1 + \beta_2 \cdot (\frac{1 - e^{-\tau\lambda}}{\tau\lambda}) + \beta_3 \cdot (\frac{1 - e^{-\tau\lambda}}{\tau\lambda} - e^{-\lambda\tau}) - \frac{C(\tau)}{\tau} \tag{3.7}$$

with,

$$\frac{C(\tau)}{\tau} = \frac{1}{2\tau} \sum_{j=1}^{3} \int_{\tau} (\Sigma' B(u) B(u)' \Sigma)_{j,j} du \tag{3.8}$$

Where $C(\tau)$ and $B(\tau)$ are solutions to the corresponding system of ordinary differential equations.

Both in-sample and out-of-sample, the AFNS performs well, outperforming most benchmark models. However, when specifically assessing out-of-sample forecasting performance in comparison to the standard DNS model, the increased complexity of the AFNS model raises questions about its relative effectiveness. A study by Coroneo et al. (2011) examined the incremental value of the AFNS model compared to the DNS model. They found that the factor loadings under the AFNS model are not statistically different from the loadings of the DNS model at a 95% confidence level. Furthermore, they found that the results from the AFNS do not significantly differ from the performance of the factors from the DNS model, suggesting that the DNS model is compatible with the no-arbitrage constraints.

3.4.3 Time-Varying Parameters

Koopman et al. (2010) introduced the time-varying parameters in the DNS model by treating λ as the fourth latent factor and allowing λ and the volatility of the interest to vary over time. Specifically, Koopman et al. proposed that λ will follow an auto-regressive process similar to the other three latent factors and introduced a common GARCH volatility component. These extensions significantly improved the model fit compared to the original DNS model, with the GARCH extension demonstrating the most substantial improvement. Notably, allowing λ to vary over time makes the model nonlinear. Due to the non-linearity of the extended model, the Kalman filter is unsuitable for estimation when the model is in state-space form. The extended Kalman filter is used instead.

4 Methodology

This section describes the core model utilized in this study, elaborating on its structure and the incorporation of macroeconomic variables. As the study progresses, various extensions to this model will be introduced and explored. Furthermore, this section describes the methodology employed in the model's estimation process. It concludes by outlining the approaches for comparing the performance across different models and conducting scenario analysis with the established model.

4.1 Model Specifications

In this study, the central model employed is an augmented version of the DNS model, wherein macro variables are integrated into the state-space framework. To maintain parsimony in the model, only three fundamental macro variables are incorporated in the baseline model, following the work of Diebold and Li (2006). In addition, we assume that factors in the transition equation evolve according to a VAR of order one. The comprehensive representation of the DNS model in state-space form, incorporating three macro variables denoted as MV_t, at time $t = 1, 2, ..., T$ is given by

$$\underbrace{\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \\ MV1_t \\ MV2_t \\ MV3_t \end{pmatrix}}_{\text{observation vector}} = \underbrace{\begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} & 0 & \dots & 0 \\ 1 & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} - e^{-\tau_N\lambda} & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & 0 \\ \vdots & & \vdots & 0 & 1 & 0 \\ 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{measurement matrix}} \underbrace{\begin{pmatrix} L_t \\ S_t \\ C_t \\ MV1_t \\ MV2_t \\ MV3_t \end{pmatrix}}_{\text{state vector}} + \underbrace{\begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{\text{residuals}}, \tag{4.1}$$

$$\underbrace{\begin{pmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \\ MV1_t - \mu_{MV1} \\ MV2_t - \mu_{MV2} \\ MV3_t - \mu_{MV3} \end{pmatrix}}_{\text{state vector demeaned}} = \underbrace{\begin{pmatrix} \phi_{11} & \dots & \dots & \phi_{16} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \phi_{61} & \dots & \dots & \phi_{66} \end{pmatrix}}_{\text{transition matrix}} \underbrace{\begin{pmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \\ MV1_{t-1} - \mu_{MV1} \\ MV2_{t-1} - \mu_{MV2} \\ MV3_{t-1} - \mu_{MV3} \end{pmatrix}}_{\text{lagged state vector demeaned}} + \underbrace{\begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ \eta_{6,t} \end{pmatrix}}_{\text{residuals}}. \tag{4.2}$$

Where equation 4.1 and 4.2 correspond to the measurement equitation and transition equation, respectively. The equivalent vector-matrix corresponds to the vector-matrix notation in equation 3.5 and equation 3.6 respectively. For convenience, they are restated below.

$$y_t = \Lambda f_t + \epsilon_t.$$
$$(f_t - \mu) = \Phi(f_{t-1} - \mu) + \eta_t,$$

For the Kalman filter to achieve optimality within the context of linear least squares estimation, it is necessary that the white noise components associated with transition and measurement disturbances exhibit orthogonality relative to one another and to the initial state of the system:

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim WN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H & 0 \\ 0 & Q \end{pmatrix} \right], \tag{4.3}$$

$$E(f_0 \epsilon_t') = 0, \tag{4.4}$$

$$E(f_0 \eta_t') = 0. \tag{4.5}$$

Moreover, the H matrix is assumed to be diagonal and the Q matrix is assumed to be unrestricted. A diagonal H matrix implies that deviations in the observed yields are uncorrelated across the different maturities. This assumption enhance computational tractability, due to the significantly reduction in the number of parameters that have to be estimated. In addition an unrestricted Q matrix allow for correlated shocks to the factors.

$$H = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \sigma_N & 0 & 0 & 0 \\ 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ 0 & \dots & \dots & 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} & q_{15} & q_{16} \\ q_{12} & q_{22} & q_{23} & q_{24} & q_{25} & q_{26} \\ q_{13} & q_{23} & q_{33} & q_{34} & q_{35} & q_{36} \\ q_{14} & q_{24} & q_{34} & q_{44} & q_{45} & q_{46} \\ q_{15} & q_{25} & q_{35} & q_{45} & q_{55} & q_{56} \\ q_{16} & q_{26} & q_{36} & q_{46} & q_{56} & q_{66} \end{pmatrix}. \tag{4.6}$$

As shown above, incorporating macroeconomic factors into the Yield-Only model is a straightforward extension. It is noticeable that the macro factors do not affect the yield curve directly. The macro factors influence the term structure indirectly via the interactions with the latent factors in the transition equation. This is done by setting the loadings on the macro variables to zero in the measurement equation. In addition, we assume that the macro variables are perfectly observed, causing the error term in the measurement equation to be equal to zero.

4.2 Kalman filter and estimation

The Kalman filter, introduced by Kalman (1960), is a two-stage recursive procedure designed to construct the likelihood function of a state vector of a dynamic system from a series of noisy measurements. This filter is applied in various fields, such as signal processing and time-series analysis. The filter is particularly effective when the system is linear and the errors in the measurements follow a Gaussian distribution. Given these assumptions and the recursive nature of the Kalman filter, it provides optimal estimates of the state vector by iteratively incorporating new measurements and refining predictions. To show the steps taken for the context of the DNS model, first equation 4.1 and 4.2 are rewritten to

$$y_t = \Lambda f_t + \epsilon_t$$

$$f_t = \tilde{\mu} + \Phi f_{t-1} + \eta_t,$$

where $\tilde{\mu}$ is equal to $(I - \Phi)\mu$. In addition, assume that the disturbance of the measurement and transition equation are mutually uncorrelated Gaussian. Then, some notations are set, after which

the predicting and updating stage, initialization, and construction of the likelihood of the Kalman filter is explained.

$\mathcal{Y}_t$ represent the set of observations $\{y_0, ..., y_{t-1}, y_t\}$ which contain the information up to time t. This filter objective is to estimate the latent state vector f_t. For now, assume the parameters of the system (Λ, Φ, Q, H, and μ) are known. Define $E[f_{t-1}|\mathcal{Y}_{t-1}] = \hat{f}_{t-1|t-1}$ and $E[f_t|\mathcal{Y}_{t-1}] = \hat{f}_{t|t-1}$, which represent the conditional expectation of the state vector at time t and $t-1$ conditional upon the information set $\mathcal{Y}_{t-1}$. The associated conditional variance of the state vector at time t and $t-1$ is defined as $V(f_t|\mathcal{Y}_t) = P_{t|t}$ and $V(f_t|\mathcal{Y}_{t-1}) = P_{t|t-1}$, respectively. Let $E[y_t|\mathcal{Y}_{t-1}] = \hat{y}_{t|t-1}$, representing the forecast of y_t conditional upon the information set with associated variance F_t. And finally, define $v_t = y_t - \hat{y}_{t-1}$ as the prediction error.

Prediction step: In the prediction step, the estimation for the state vector in the upcoming period, $\hat{f}_{t|t-1}$, relies on the updated estimate of the factor from the preceding period ($\hat{f}_{t-1|t-1}$). This process generates the variance of the state vector as a by-product ($P_{t|t-1}$), which is subsequently employed in subsequent stages to refine the estimation of the state vector. The prediction steps are given by

$$\hat{f}_{t|t-1} = \tilde{\mu} + \Phi \hat{f}_{t-1|t-1} \tag{4.7}$$

$$P_{t|t-1} = \Phi P_{t-1|t-1} \Phi' + Q \tag{4.8}$$

Update/ Filter step: Once the new observation y_t comes in, the current estimate of the state vector can be updated($\hat{f}_{t|t}$) by taking a weighted average of the initial estimation and the prediction error v_t. The Kalman gain, which is given by $K_t = P_{t|t-1}\Lambda' F_t^{-1}$, determines how much weight is given to the prediction error. In updating the state estimate, less weight is assigned to the prediction error v_t, The larger the prediction error variance (F_t). The update steps are given by

$$\hat{f}_{t|t} = \hat{f}_{t|t-1} + K_t v_t \tag{4.9}$$

$$P_{t|t} = P_{t|t-1} - K_t \Lambda' P_{t|t-1} \tag{4.10}$$

Initialization: The unconditional mean and variance of the state vector are used to initialize the Kalman filter. These are suitable choices since the state vector is often assumed to be stationary.

Likelihood Construction: Given the parameter values of matrices Λ, Φ, Q, H, and μ, the Kalman filter facilitates the extraction of latent factors following the above procedure. Let θ be a vector containing all the parameters of these matrices. In the absence of explicit parameter values, the Kalman filter constructs the joint conditional probability distribution function, $f(y_1, ..., y_T|\theta)$, for the state vectors, assuming a jointly normal distribution under the normality assumptions of the disturbance terms. It can be shown that the log-likelihood function, denoted as $l(\theta)$, can be given by equation 4.11. Maximizing the log-likelihood over θ involves finding parameter values that optimize the likelihood of the observed data arising from the model with the estimated parameters.

$$l(\theta) = -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T} v_t' F_t^{-1} v_t \tag{4.11}$$

4.3 Model comparison

In this study, various extensions of the DNS model will be evaluated for both the US and the EA. These adaptations will be explored in the context of modeling and forecasting Treasury yields in

the US and swap rates in the EA. With the aim of conducting scenario analysis using these models, it is essential to assess and compare the in-sample fitting capabilities and out-of-sample forecasting accuracy of each model variation. The in-sample fitting capabilities are assessed based on the mean and standard deviation of the errors, the RMSEs, and the Mean Absolute Error (MAE). To facilitate forecasting accuracy, the out-of-sample RMSEs and the average overall RMSE of all maturities will be employed as the primary measure for evaluating the performance of these extended DNS models. Averaging over all RMSEs captures the performance of a model in a single number. In addition to averaging the RMSE, the MAE per model is assessed. The RMSE gives more weight to large errors, while the MAE gives a straightforward average of absolute errors, giving a more direct interpretation of the magnitude of the average error. If the average RMSE is significantly bigger than the MAE, it implies the presence of large errors. The model exhibiting the lowest average RMSE is regarded as the optimal choice for conducting scenario analysis.

A formal description of the calculation for each out-of-sample error metric is presented, where $\hat{y}(\tau_i)_{T+h}$ denotes the estimated interest rate for maturity τ_i at time $T+h$, $y(\tau_i)_{T+h}$ represents the realized interest rate for maturity τ_i at time $T+h$, H denotes the forecast horizon in months, and N indicates the number of different maturities.

$$RMSE(\tau_i) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{y}(\tau_i)_{T+h} - y(\tau_i)_{T+h})^2} \tag{4.12}$$

$$Average\ RMSE(\tau_i) = \frac{1}{N}\sum_{i=1}^{N} RMSE(\tau_i) \tag{4.13}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\sum_{h=1}^{H} |\hat{y}(\tau_i)_{T+h} - y(\tau_i)_{T+h})^2| \tag{4.14}$$

The predictive performance of each model is assessed based on three different forecasting horizons: one month, one year, and five years. The forecast performance of each extension is benchmarked against the forecasting performance of the random walk model. An overview of the models and models extensions are given in table 1

Label	Model Description
RW	Random Walk
YO-OLS	2-step estimation of Yield-Only model
YO-KF	1-step estimation of Yield-Only model
YM3-KF	1-step estimation of Yield-Macro model with 3 macro variables
YM4-KF	1-step estimation of Yield-Macro model with 4 macro variables
YO-OLS-diff	2-step estimation of Yield-Only model with stationary factors
YO-KF-diff	1-step estimation of Yield-Only model with stationary factors
YM3-KF-diff	1-step estimation of Yield-Macro model with 3 macro variables & stationary factors
YM4-KF-diff	1-step estimation of Yield-Macro model with 4 macro variables & stationary factors

Table 1: Overview of different models

4.4 Scenario Analysis

Scenario analysis is conducted using the model that exhibits the minimum out-of-sample error metric. The following steps are adopted post-identification of the optimal model:

1. **Main Scenario Development**: Employ the central bank's forward guidance and projections of macroeconomic variables as the main scenario. To align the data frequency with the model's requirements, interpolate these macro variables to obtain monthly series.

   (a) **Alternative Scenario Construction**: Generate additional scenarios by formulating alternative projections for the macroeconomic variables based on plausible economic developments or policy changes.

2. **Recursive Determination of Latent Factors**: Utilize the transition equation of the selected model to recursively calculate the latent factors level, slope, and curvature. It is imperative to note that while the latent factors evolve according to the transition equation, the macroeconomic variables are treated as exogenous, evolving according to their scenario-specified paths.

3. **Term Structure Projection**: Once the latent factors are determined for the specific scenarios, the measurement equation is used to project the term structure. This step incorporates the estimated decay parameter, ensuring that the curve's evolution reflects the underlying dynamics of the model and the specified scenario.

By following these structured steps, the scenario analysis yields projections of the term structure under different economic conditions, reflecting both the central bank's expectations and alternative economic narratives. This approach provides a comprehensive view of potential future term structure movements

5 Explanatory Data Analysis

5.1 Treasury yields

The monthly Treasury data from January 1986 through December 2023, sourced from the U.S. Department of Treasury database, was utilized for this analysis. The dataset comprises yields with various maturities measured in months, including 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, 204, 216, 228, 240, 252, 264, 276, 288, 300, 312, 324, 336, 348, and 360. Figure 4 illustrates the dynamics of the Treasury yield curve from January 2000 through November 2023 using a three-dimensional (3D) plot. To enhance clarity and conciseness, not all years are displayed in this graph. Table 3 presents the descriptive statistics of the most relevant yields and the economic interpretation of the three underlying factors. The average yield curve exhibits an upward slope, indicating higher yields for longer-term maturities. A graph depicting the average yield curve and the 25% and 75% quantiles can be found in Appendix A.1. Moreover, short-term yields demonstrate greater volatility compared to longer-term yields. Additionally, the persistence of long-term yields, as measured by its autocorrelation, is higher than that of short-term yields. The three latent factors, measured as level, slope, and curvature, exhibit distinct characteristics. The level factor displays the highest degree of persistence, signifying a sustained influence on the yield curve. Conversely, the curvature factor exhibits lower persistence but is significantly more volatile relative to its mean compared to the other factors.

5.2 Euribor swap rates

The Euro area zero-coupon yield curves, as published on the ECB's website, is a representative index and is composed of triple-A rated bonds issued in euros by the national governments of member states within the EA. These averages reflect the absence of a single sovereign bond issuer in the EA in contrast to the US Treasury. Instead, each member state is responsible for issuing its own government bonds in the EA. Consequently, the yield curve for each country inherently represents its specific credit risk, economic conditions, and monetary policy stance. Given the diverse and multi-national nature of the Euro area's bond market, the 6-month Euribor swap rate is often utilized as an analog to the US Treasury yields. Euribor swap rates offer a uniform benchmark across the Euro area, reflecting a uniform rate market participants use for a wide range of financial products and contracts. In this study, the Euribor swap rate represents the fixed rate of the swap contract with maturities ranging from 6 months to 30 years, and the 6-month Euribor interest rate represents the floating rate of the swap contract. Figure 5 illustrates the dynamics of the 6-month Euribor swap rate from January 2002 through December 2023. The evolution of the Euribor swap rates closely resembles the evolution of Treasury yields. Also, the descriptive statistics show a similar pattern as in the US, with an upward-sloping curve and persistency increasing as a function of maturity reflected by the autocorrelation and standard deviation. On average, the Euribor swap rate is lower for each maturity and has a higher standard deviation relative to its mean. Noticeably, Euribor swap rates reached negative territory. This can be addressed in the different monetary policies, where the ECB employed a negative deposit facility rate. In contrast, the Fed kept its policy rate above zero. The empirical factors show the same characteristics and interrelations as the empirical factors of the US.

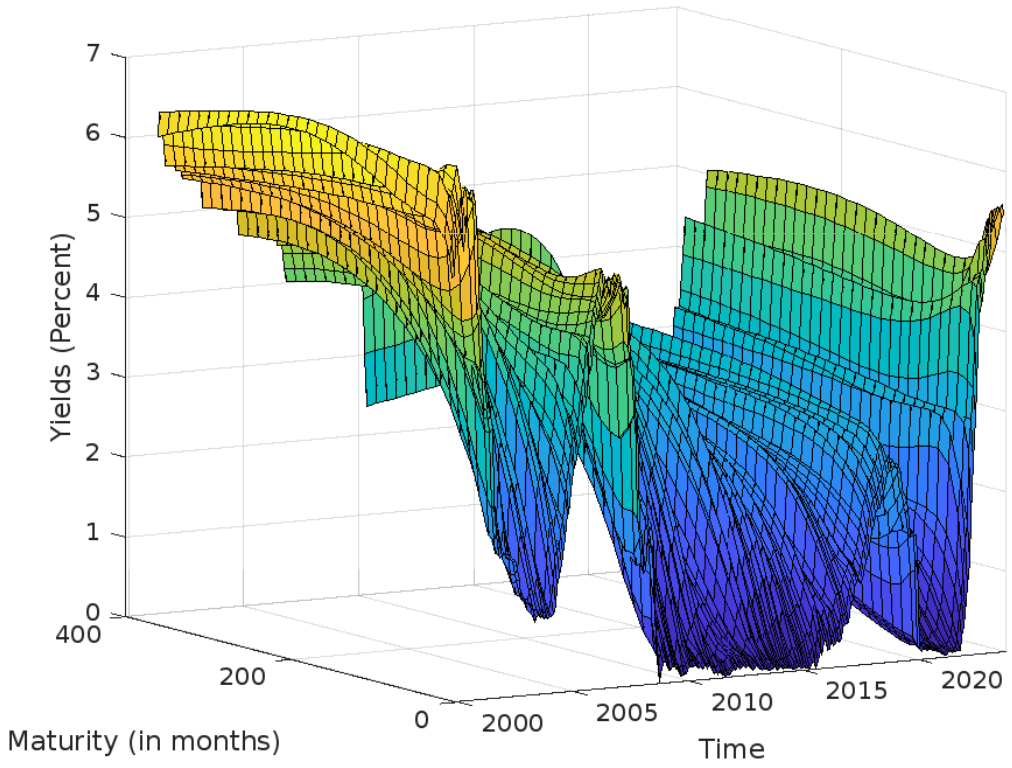


Figure 4: 3D plot for the US Treasury yield curves

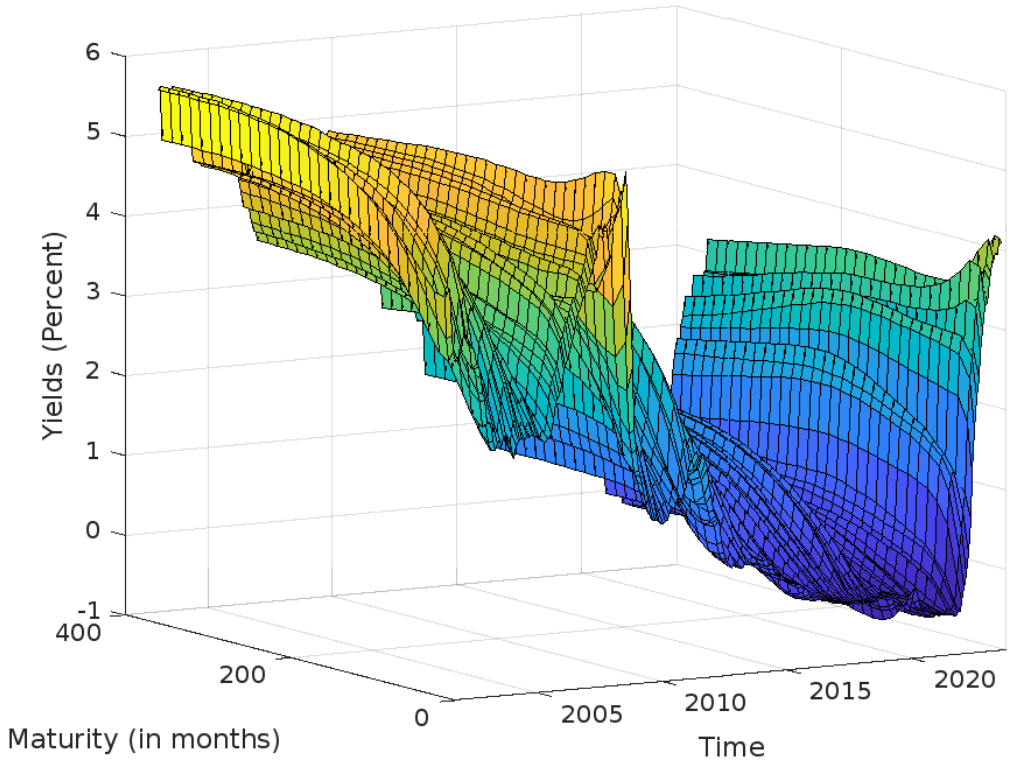


Figure 5: 3D plot for the EA 6-month Euribor swap curves

Maturity (Months)	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	3.099	2.546	0.000	9.230	0.992	0.852	0.487
12	3.406	2.608	0.068	9.658	0.992	0.872	0.538
24	3.626	2.596	0.113	9.566	0.991	0.888	0.611
60	4.190	2.443	0.243	9.317	0.989	0.907	0.723
120	4.817	2.297	0.552	9.642	0.989	0.916	0.773
240	5.273	2.177	1.036	9.956	0.988	0.921	0.791
360	5.281	2.077	1.383	10.160	0.986	0.914	0.786
Level	5.281	2.077	1.383	10.160	0.986	0.914	0.786
Slope	1.718	1.288	-1.835	4.375	0.967	0.545	-0.134
Curvature	-0.664	0.898	-2.691	1.749	0.947	0.611	0.150

Table 2: Descriptive Statistics, US yield curves

Maturity (Months)	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
6	1.375	1.643	-0.541	5.405	0.988	0.731	0.390
12	1.551	1.673	-0.570	5.365	0.987	0.774	0.473
24	1.676	1.679	-0.579	5.314	0.987	0.800	0.523
60	2.043	1.681	-0.493	5.194	0.987	0.834	0.581
120	2.491	1.649	-0.248	5.500	0.987	0.852	0.596
240	2.796	1.619	-0.019	5.721	0.988	0.863	0.606
360	2.773	1.615	-0.054	5.715	0.987	0.865	0.611
Level	2.773	1.615	-0.054	5.715	0.987	0.865	0.611
Slope	1.398	0.940	-1.367	2.953	0.961	0.416	-0.097
Curvature	-1.046	0.785	-2.392	1.624	0.946	0.534	0.058

Table 3: Descriptive Statistics, 6-month Euribor swap rate curves

5.3 Macro variables

To characterize the relationship between latent factors and the macroeconomy, we adopt the approach of Diebold et al. (2006) and employ three fundamental macro variables as essential indicators for both the US and the EA: Economic activity, central bank policy rate, and annual core price inflation. In the US, manufacturing capacity utilization (CU_t), measured as manufacturing output relative to its potential, is used to measure real economic activity. In the EA, GDP_t is used as a measure of economic activity. In the respective models for the US and EA, key policy interest rates are represented by the upper limit of the Federal Funds Rate target range (FFR_t) for the US, set by the Federal Reserve, and the Main Refinancing Operations rate (MRO_t) for the EA, set by the ECB. The Harmonised Index of Consumer Prices ($HICP_t$) is employed as the standard measure for core price inflation in the European model, ensuring consistency and comparability across EA member states. Conversely, the Personal Consumption Expenditures Price Index (PCE_t) is utilized in the US model, reflecting its status as the preferred inflation metric by the Federal Reserve. These variables are selected as the minimum set necessary to capture the fundamental dynamics of macroeconomic activity, capturing the essence of monetary policy and inflationary trends within the respective economies. The descriptive statistics for the US and the EA are given in table 4 and 5, respectively. The corresponding variables in the US and the EA display similar dynamics, with

Macro variables	Mean	Std. dev	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
CU	77.684	4.167	62.249	85.572	0.980	0.761	0.344
FFR	3.307	2.641	0.250	9.813	0.993	0.854	0.473
Core PCE	2.311	1.078	0.626	5.575	0.986	0.761	0.312
BSP	12.806	5.967	4.420	26.063	0.997	0.918	0.686

Table 4: Descriptive statistics, macro variables US

Macro variables	Mean	Std. dev	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
GDP	3.168	3.146	-11.635	15.767	0.965	0.209	-0.203
MRO rate	1.339	1.401	0.000	4.500	0.981	0.685	0.333
Core HICP	1.617	1.044	0.210	5.663	0.971	0.540	-0.076
EL	0.090	1.034	-0.705	2.715	0.990	0.870	0.484

Table 5: Descriptive statistics, macro variables EA

the policy rate being the most persistent factor. On average, all variables are positive.

6 Measure for Balance Sheet Policy

In this section, different methods to quantify the unconventional monetary policies for each of the central banks are introduced. The methods are tested against each other, and their performance is assessed based on out-of-sample RMSE.

6.1 US

To quantify the balance sheet policy of the Fed, the following ratio is used as, following the work of Loonen (2022) :

$$BSP_t = \frac{TR_t^{Fed} + MBS_t^{Fed}}{TR_t^{market} + MBS_t^{market}} \tag{6.1}$$

In the numerator, the variables TR_t^{Fed} and MBS_t^{Fed} represent the value of Treasury securities and MBSs held by the Federal Reserve at time t, respectively. In the denominator, the variables TR_t^{market} and MBS_t^{market} represent the total value of Treasury securities and MBSs outstanding in the market at time t, respectively. The resulting ratio provides a measure of the scale of the Fed's operations relative to the overall size of the market. As demonstrated by Loonen (2022), this ratio, within the context of the YM4-KF model, exhibited superior out-of-sample Root Mean Square Error (RMSE) performance compared to alternative measures such as the standardized total of the Fed's balance sheet or the net supply of public long-term debt as defined by Gagnon, Raskin, Remache, and Sack (2011). Moreover, an increase in the lead time before the implementation of QE policies has been shown to decrease the out-of-sample RMSE, indicating a predictive benefit to incorporating the market's perception into the model. Therefore, in alignment with these findings, a lead time of six months is adopted in the present study.

6.2 EA

Given the distinct structure, prevailing economic conditions, and policy implementation approaches within the EA, the following two variables are considered to measure the balance sheet policy of the ECB, diverging from the Balance Sheet Policy (BSP) ratio utilized in the U.S. model:

1. Excess liquidity in the Euro banking system
2. Balance sheet to GDP ratio

Excess liquidity in the Euro banking system is defined as the surplus of money deposited at the national banks after commercial banks have met their required minimum level of reserves. Excess liquidity in the Euro system is a product of the unconventional monetary policies employed by the ECB. After the financial crisis of 2008, there was less trust and willingness across banks to lend money to each other, making the ECB the only reliable source of liquidity. During this time, banks demanded more liquidity than needed to meet their short-term liabilities, and the ECB provided as much liquidity as demanded by banks, leading to excess liquidity in the system. Due to the APP, PEPP, and TLTRO, the excess of liquidity has grown further. The interest paid on this surplus of liquidity stored at the deposit facility is called the deposit facility rate. The ECB is the source of this excess liquidity; therefore, excess liquidity reflects the size and expansion of the central bank's balance sheet. Figure 6 displays the amount of excess liquidity (orange), calculated as the difference between the liquidity-providing factors (dark blue) and the liquidity-absorbing factors (light blue). The factors that provide liquidity are the ECB's main refinancing operations (MRO),

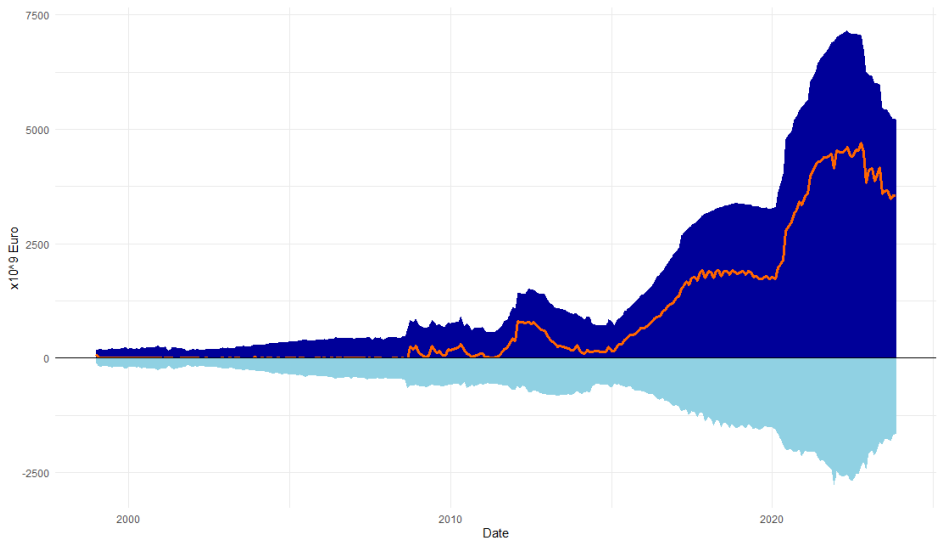


Figure 6: EA excess liquidity

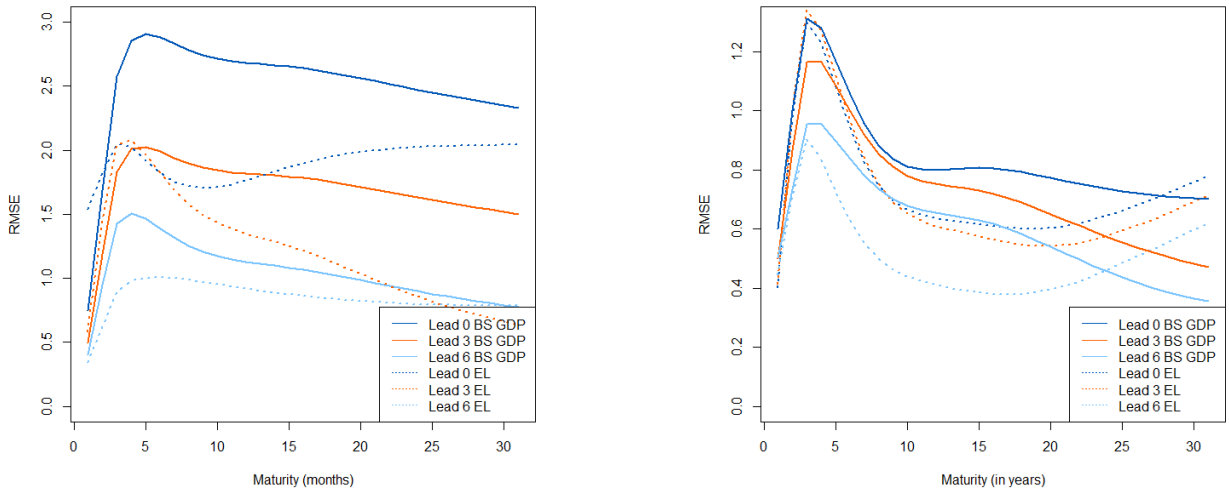
(targeted) longer-term refinancing operations (TLTRO), marginal lending facilities (MLF), and the securities held for monetary policy purposes (SMP). The sum of these factors is represented by the blue graph. ECB's deposit facility (DF) and the minimum required reserves (MRR) are considered liquidity-absorbing factors. The total effect of the liquidity absorbing factors is determined as the sum of the deposit facility and the actual outstanding reserves (AUR), presented by the light blue graph. Equation 6.2 depicts the calculation of excess liquidity in the euro banking system. All information about the liquidity-providing and absorbing factors is derived from the balance sheet of the ECB.

$$EL_t = MRO_t + TLTRO_t + MLF_t + SMP_t - (DF_t + MRR_t + (AUR_t - MRR_t)) \tag{6.2}$$

The balance sheet to GDP ratio measures the size of the balance sheet of the ECB relative to the GDP of the EA. This ratio is an indicator of the ECB's monetary stance. A high ratio suggests a looser monetary policy, reflected by liquidity injections in the economy in the form of QE. A low ratio indicates a tightened monetary policy stance characterized by liquidity absorption in the form of QT.

Each variable is standardized by adjusting for its mean and dividing by its standard deviation. This standardization procedure transforms variables to a common scale, allowing for direct comparisons between them. It also compensates for differences in the magnitudes and units of the other macro variables in the model, ensuring that its substantial magnitude does not disproportionately skew the analysis. The performance of these standardized variables was subsequently evaluated based on the RMSE within the YM4-KF model framework across three different leads. The lead times in months are 0, 3, and 6, categorized as short-term, medium-term, and long-term, respectively. QE announcements typically precede their implementation by a period ranging from several days to months. The lead times are ranging from days to several months before implementation. The specific lead time is influenced by factors such as economic urgency, prevailing market conditions, and the strategic objectives of the monetary policy. The change in yield following a QE announcement reflects the investor's expectations and is largely due to market pricing in expected future purchases. The model is trained and tested based on the 70%-30% principle, and two data

samples are used. Figure 7a displays the out-of-sample RMSE trained and tested on all available data, ranging from 2002 to the present. Figure 7b displays the results based on the dataset encompassing the period from 2014, marking the first announcement and initiation of QE in the EA to the present.



(a) 2002-2023

(b) 2014-2023

Figure 7: RMSE BSP variable

A wide spectrum of maturities was targeted under the APP and PEPP to different segments of the term structure, focusing on mid to long-range maturities. This strategy is evidenced by the graphical representations provided, particularly noting a consistent reduction in Root Mean Square Error (RMSE) beyond the five-year maturity mark. Comparative analysis indicates that the YM4 model, incorporating excess liquidity as the fourth macro-variable, consistently outperforms its counterpart, employing the balance sheet-to-GDP ratio based on RMSE across both data samples. Additionally, incorporating the announcement effect universally reduces RMSE across all models, underscoring its significance in predictive accuracy. A notable observation across the models is the relatively stable RMSE at the shorter end of the term structure, regardless of the specific fourth macro-variable employed or the consideration of lead times. This suggests that both excess liquidity and balance sheet-to-GDP ratio, alongside lead time incorporation, effectively capture the dynamics predominantly in the mid and long-term spectrum of the term structure. Interestingly, models trained on the shorter data sample exhibit lower RMSE, potentially indicative of regime shifts associated with adopting unconventional monetary policies. Furthermore, both models demonstrate similar RMSE patterns, which can be attributed to the high correlation (0.967) observed between the two variables. This correlation underscores the intertwined nature of excess liquidity and balance sheet metrics in reflecting underlying economic and monetary conditions.

7 Performance four macro model

In this section, a detailed evaluation of the model incorporating the central bank's balance sheet is presented. The performance of the YM4-KF is compared to the baseline models and the random walk, examining both in-sample and out-of-sample predictions across different metrics. The out-of-sample forecasting accuracy is examined across different forecast horizons to asses the model's predictive reliability over time. Furthermore, the fitting capabilities of the models is assessed by fitting the average interest rate curve, complemented by a visualization of the interpretation of the latent factors.

7.1 In-sample fit

The in-sample fit of each model is evaluated through an examination of descriptive statistics for sample errors across various maturities. These statistics are detailed in Table 6, providing insights into the short-term, intermediate-term, and long-term aspects of the term structure. The chosen maturities (3/6 months, and 1, 5, 10, 20, 30 years) are considered benchmark maturities and actively traded. The mean of each descriptive statistic is calculated across all maturities. The minimum value for each metric at each maturity is highlighted in bold, indicating the superior-fitting model for that specific maturity. Additionally, the minimum mean error metric is bolded, indicating that the particular model has the best overall in-sample fit.

The mean error provides a somewhat distorted representation as negative and positive errors offset each other, resulting in a lower mean error. MAE and RMSE are included to offer a more comprehensive view of error terms. All models exhibit a remarkable fit to the respective term structures based on MAE and RMSE. Comparison of means for each descriptive statistic reveals subtle disparities, indicating minimal deviations between the realized rates and the fitted models. The YM4-KF outperforms other models in both the US Treasury yields and EA swap rates based on standard deviation, MAE, and RMSE. However, the differences between competing models are small. The inclusion of central bank balance sheet information marginally enhances in-sample fit compared to the YM3-KF and YO-KF models. Additionally, the two-step and one-step approaches show minimal discrepancies in overall fit, with MAE ranging between 0.001 and 37.712 basis points for EA swap rates and between 1.629 and 29.051 basis points for the Treasury yields. Across different maturities, models exhibit their largest sample errors at the front end (3/6 month rate) and at the back end (30 years rate) of the term structure, with significantly larger errors at the front end. This characteristic is consistent across all models in both data samples. Models estimated with the one-step approach closely fit the data for the mid-term maturities, surpassing those estimated with the two-step approach. Conversely, models based on the two-step approach fit the data less accurately at the short and long ends of the term structure. Figures 8 and 9 visually represent the average in-sample fit of the YM4-KF and YO-OLS models onto the average term structure in the US and EA, respectively, excluding the 3-month yield and 6-month swap rate. These graphs illustrate that the YM4-KF model fits the data well for maturities up to 20 years but starts deviating and overestimating the data beyond this maturity. The YO-OLS model underestimates the maturities up to 5 years. After this point, the steepness of the graph declines, letting it overestimate the rates with maturity between 5 and 12 years and beyond 25 years.

Figure 10 showcases the latent factors derived from the YM4-KF model, identified as the best in-sample performer, represented in black. The empirical counterparts are illustrated in blue. The Kalman smoother is used to obtain optimal extraction of the latent factors. The observed correla-

tion between the estimated factors and their empirical counterparts are found to be $0.995, 0.977$,and 0.930, reinforcing the interpretation of the model's latent factors as level, slope, and curvature. A similar graphical representation of the YM4-KF model applied to the US is available in Appendix A.2, where the correlation between the latent factors and their empirical counterparts equals $0.986, 0.977$, and 0.878, for level, slope, and curvature, respectively.

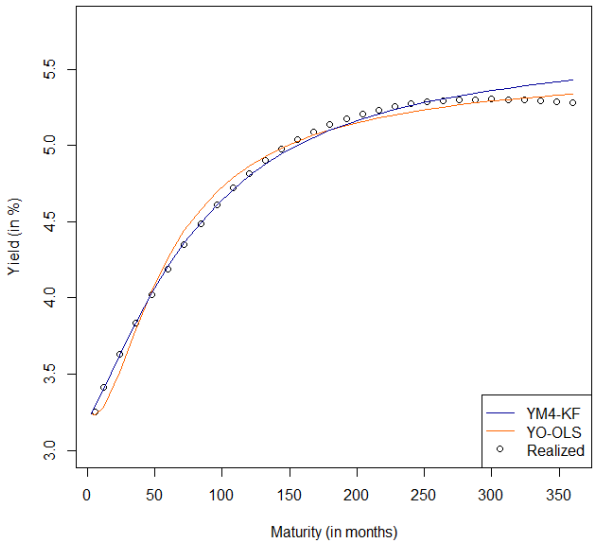


Figure 8: In-sample fit of the Treasury yield curve in the US

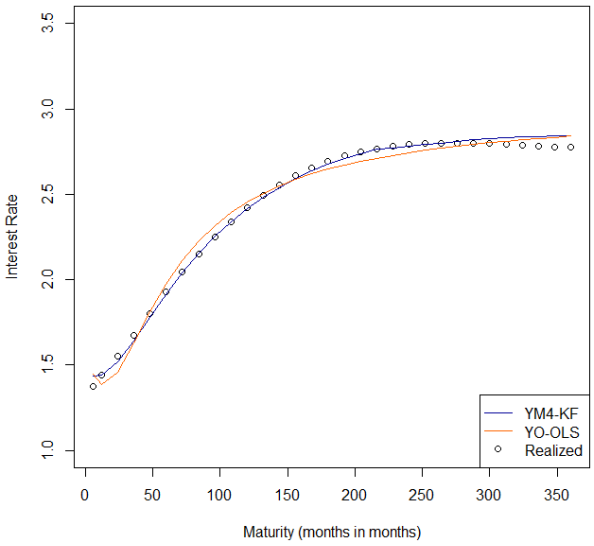


Figure 9: In-sample fit of the average swap curve in the EA

	YO-OLS	YO-KF	YM3-KF	YM4-KF	YO-OLS	YO-KF	YM3-KF	YM4-KF
	Mean Error				Mean Error			
3/6m	**-12.645**	-14.251	-13.496	-13.583	-6.964	**-4.827**	-5.967	-5.966
1	13.259	**0.480**	0.919	0.871	5.291	-1.181	**0.000**	**0.000**
5	-7.254	**-1.410**	-1.414	-1.413	-4.821	**-0.856**	1.237	1.233
10	-4.910	**1.694**	1.702	1.701	-3.611	1.570	0.242	**0.240**
20	5.232	**1.095**	1.096	**1.095**	4.694	**1.310**	-2.313	-2.312
30	**-5.862**	-14.741	-14.736	-14.734	**-6.607**	-13.741	-18.158	-18.156
mean	**0.000**	-1.856	-1.739	-1.702	**0.000**	-1.286	-3.184	-3.185
	Standard deviation				Standard deviation			
3/6m	**10.451**	32.307	31.858	31.898	**6.529**	47.064	16.061	16.059
1	10.760	**1.026**	2.257	2.146	5.154	27.547	0.004	**0.002**
5	**6.081**	11.087	11.050	11.022	**2.722**	3.184	5.123	5.117
10	10.730	5.058	5.053	**5.001**	5.685	2.249	1.797	**1.796**
20	5.337	1.898	1.899	**1.855**	1.694	**1.116**	2.172	2.169
30	**19.675**	22.306	22.324	22.315	8.844	**7.944**	11.697	11.697
mean	9.337	9.168	9.180	**9.032**	5.026	5.561	**4.837**	**4.837**
	MAE				MAE			
3/6m	**13.902**	29.051	28.529	28.590	**8.326**	37.712	12.803	12.800
1	14.574	**0.812**	1.793	1.708	6.392	22.318	0.003	**0.001**
5	**7.556**	7.836	7.825	7.830	4.849	**2.546**	4.415	4.410
10	9.589	3.664	3.664	**3.663**	5.839	1.912	1.327	**1.330**
20	5.910	1.629	**1.630**	**1.630**	4.694	**1.380**	2.378	2.378
30	**15.639**	20.756	20.757	20.756	**7.361**	13.741	18.158	18.156
mean	8.678	7.666	7.673	**7.552**	5.579	5.919	5.580	**5.456**
	RMSE				RMSE			
3/6m	**16.398**	35.269	34.567	34.637	**9.538**	47.223	34.567	17.103
1	17.068	**1.146**	2.435	2.314	7.380	27.521	2.435	**0.002**
5	**9.461**	11.163	11.128	11.136	5.534	**3.291**	11.128	5.254
10	11.790	5.329	5.327	**5.310**	6.726	2.739	5.327	**1.809**
20	7.470	2.239	**2.191**	**2.191**	4.989	**1.720**	2.191	3.167
30	**20.509**	26.716	26.729	26.728	**11.027**	15.865	26.729	21.586
mean	11.083	10.345	10.148	**10.049**	6.989	7.342	6.935	**6.877**

Table 6: Descriptive statistics of the in-sample error terms in the US (left panel) and the EA(right panel) in basis points. $\lambda = 0.0609, 0.0351, 0.0350, 0.0305$ for the respective models in the US, and $\lambda = 0.0609, 0.0373, 0.0372, 0.0350$ for the respective models in the EA

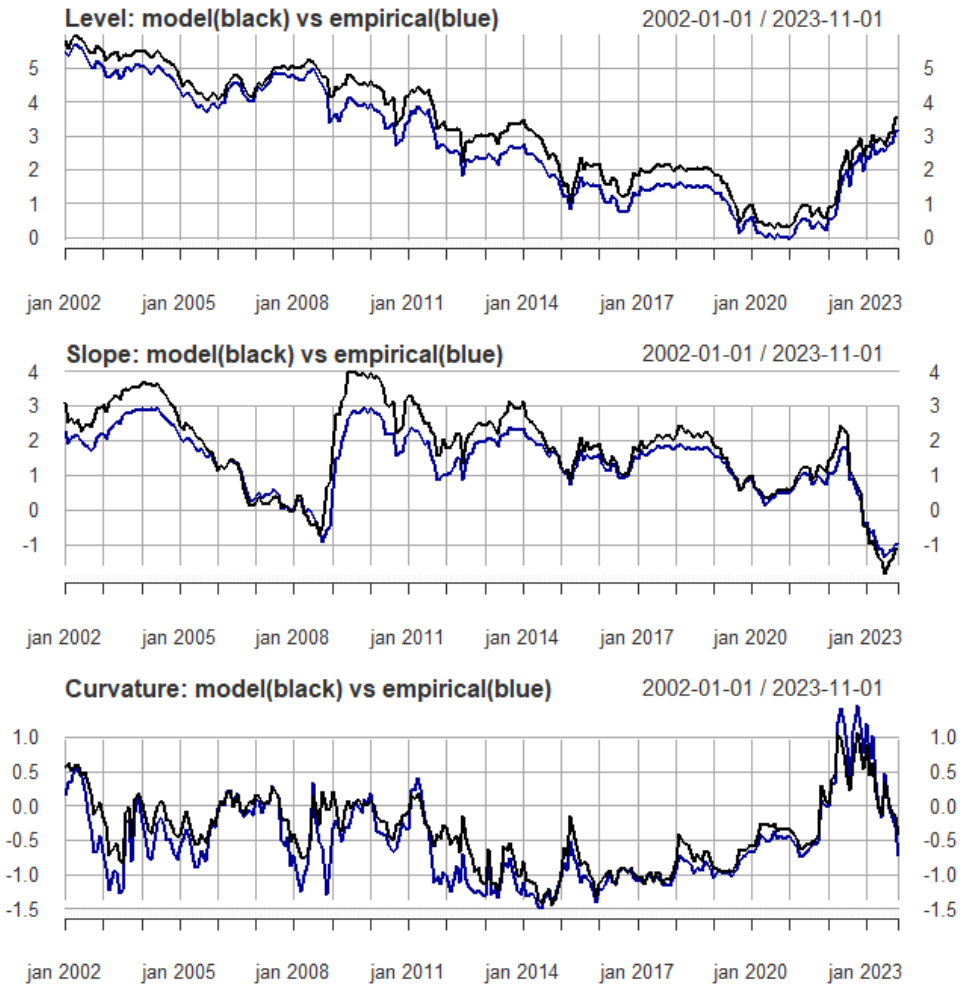


Figure 10: Estimated latent factors from YM4-KF vs empirical counterparts. Level: $\hat{L}_t$ vs $y_t(\tau_N)$. Slope: $\hat{S}_t$ vs $y_t(\tau_N) - y_t(\tau_1)$. Curvature: $0.5\hat{C}_t$ vs $y_t\tau_5 - (y_t(\tau_N) - y_t(\tau_1))$

7.2 Out-of-sample forecasting performance

The forecast performance of YM4-KF is compared against YM3-KF, YO-KF and the RW (equation 7.1). The forecasting performance of the YO-OLS is added to show the improvements of the 1-step approach against the 2-step approach. Three horizons are considered to assess the forecasting performance for each model. Unlike the in-sample assessment, the out-of-sample forecast uses all the data available to determine the parameters and latent factors of the model. By employing the Nelson-Siegel framework, forecasting the interest rate curve is reduced to estimating the latent factors. Once the parameters of each model are estimated on the training data, each of the future latent factors can be determined recursively using the estimated transition matrix of the VAR(1) model. An important assumption is that the macro variables are treated as exogenous variables, implying that they evolve outside of the system.

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \epsilon_t(\tau_i), \qquad \epsilon(\tau_i) \sim \mathcal{N}(0,\, \sigma^2(\Delta y(\tau_i))) \tag{7.1}$$

Table 7 and 8 comprehensively report the out-of-sample forecasting errors evaluated with RMSE and the MAE for each of the three horizons.

7.2.1 1-month horizon

Panel A of both Table 7 and Table 8, describes the forecasting errors for the 1-month horizon within the US and the EA. The mean RMSE and MAE are similar for this horizon, as the 1-month horizon only includes a single data point for each maturity. Across both regions, no model, on average, outperforms the predictive accuracy of the random walk model for this short forecasting horizon. Additionally, the models based on the YO framework exhibit superior performance compared to models that incorporate macroeconomic variables. Within the US dataset, the YM3-KF model demonstrates the highest average performance for this horizon, with a RMSE of 0.289. The incorporation of the Federal Reserve's balance sheet into this model results in a marginal enhancement in forecasting performance, reducing the mean RMSE to 0.257. Contrastingly, in the EA dataset, the YM4-KF model exhibits the highest average forecasting error. The expectation that including the central bank's balance sheet would enhance forecasting performance, particularly for longer-term maturities, is not realized. Notably, the random walk and YO models outperform the YM4-KF for the longer-term maturities in both regions. Moreover, the analysis reveals that the one-step YO model does not significantly improve over the 2-step YO model in the US. This suggests that, for the given context and dataset, the Kalman filter does not contribute significantly to enhancing forecasting accuracy for the 1-month horizon.

7.2.2 1-year horizon

Panel B of both the aforementioned tables reports the forecasting error for a 1-year forecasting horizon. The models exhibit comparable performance in both data sets, and both tables demonstrate that forecasting accuracy consistently improves with longer forecasting horizons relative to the random walk model based on mean RMSE and MAE. The random walk model, on average, performs the poorest in both datasets, exhibiting an RMSE of 0.876 in the US and 0.928 in the EA. The best-performing model for both regions is the YM4-KF, with RMSEs of 0.508 and 0.641 for the US and EA, respectively. The YM4-KF accuracy is relatively good for the mid-term maturities ranging from 1 up to 20 years. Despite achieving the lowest mean RMSE and MAE, the YM4-KF only marginally outperforms the YM3-KF. The results are very similar for every maturity. The one-step approach consistently outperforms the two-step approach for models without macro variables in

both regions. Incorporating macro variables into the state space framework further reduces RMSE. This reduction is more pronounced for the US models than the EA models.

7.2.3 5-year horizon

In Panel C of both table 7 and table 8, the forecasting errors for the 5-year horizon are presented. The effects observed in Panel B for the 1-year horizon are amplified, highlighting the importance of considering longer forecasting horizons for model evaluation. Once again, the random walk model emerges as the worst-performing model, with consistently high RMSE values. The YM4-KF model dominates the forecasts, exhibiting the lowest overall mean RMSE and MAE. Across various maturities in both datasets, the YM4-KF consistently outperforms other models for long-term maturities, showcasing its efficiency for mid to long-term forecasting horizons. This further reinforces the model's reliability and robustness. The one-step forecasting approach continues to show improved forecast accuracy compared to the two-step approach, reinforcing the findings from previous horizons. This effect is more pronounced in the US models compared to the EA models, emphasizing the relevance of the forecasting methodology within the specific economic context. Moreover, incorporating leading macroeconomic indicators into the forecasting models significantly improves accuracy compared to the YO models. Notably, the inclusion of the central bank's balance sheet as an additional variable has a substantial impact on forecast accuracy, particularly at the back end of the term structure. The models incorporating three or four macroeconomic variables consistently outperform the models without macro variables, indicating the valuable information provided for longer forecast horizons.

7.3 Latent factors and macroeconomic impulse response functions

Short attention is given to the impulse response functions (IRF). The IRF is used to analyze the dynamic effects of shocks of the endogenous variables on the system in the EA. Recall that for forecasting and scenario analysis, the macro variables are treated as exogenous variables and evolve outside of the system. Figure 20 displays the response of the latent factors to shocks in macro variables. The IRF obtained matches the results obtained by Diebold et al. (2006). The curvature does not react much to any of the shocks in the system. The slope reacts immediately to shocks in GDP and the policy rate. A shock in GDP decreases the slope in the first instance, implying that the difference between the short and long-term rates increases. Recall that the negative estimated slope from the model is highly correlated with the empirical slope. Moreover, the slope factor increases in response to a shock in the policy rate, implying that the difference between the short and long rates decreases. Also, shocks to inflation decrease the slope factor in the short run. This behavior is consistent with the observed behavior of the term structure in response to these shocks. The level factor increases to shocks in the policy rate and inflation. The level factor initially decreases, followed by an increase in response to a shock in GDP. Both level and slope do not respond much to a shock in the excess liquidity in the EA system.

	RMSE									Mean RMSE	MAE
Maturity	3m	1	2	5	10	15	20	25	30		
Panel A: 1-month horizon											
RW	0.244	**0.072**	**0.004**	0.122	0.242	0.298	0.279	**0.198**	**0.076**	**0.199**	**0.199**
YO-OLS	0.056	0.184	0.197	**0.091**	0.179	0.300	0.331	0.282	0.182	0.222	0.222
YO-KF	0.293	0.323	0.161	0.172	0.278	0.321	**0.038**	0.301	0.222	0.217	0.217
YM3-KF	0.129	0.272	0.201	0.251	0.200	0.305	0.342	0.298	0.201	0.289	0.289
YM4-KF	**0.056**	0.099	0.184	0.113	**0.149**	**0.283**	0.332	0.297	0.205	0.257	0.257
Panel B: 1-year horizon											
RW	**0.111**	1.103	1.371	**0.421**	0.745	0.701	1.070	0.494	0.99	0.876	0.757
YO-OLS-	0.784	1.064	1.106	0.849	0.615	0.547	0.518	0.502	**0.503**	0.647	0.601
YO-KF	0.775	1.039	1.077	0.835	0.615	0.548	0.519	0.502	0.504	0.642	0.598
YM3-KF	0.836	0.370	**0.363**	0.531	0.554	0.535	0.523	0.528	0.563	0.533	0.491
YM4-K	0.760	**0.321**	0.398	0.541	**0.534**	**0.513**	**0.496**	**0.490**	0.506	**0.508**	**0.462**
Panel C: 5-year horizon											
RW	1.694	1.833	1.787	1.576	3.465	1.068	1.511	1.551	3.126	1.939	1.645
YO-OLS	1.869	1.808	1.705	1.498	1.278	1.157	1.077	1.011	0.952	1.260	1.025
YO-KF	1.835	1.758	1.621	1.372	1.176	1.076	1.012	0.96	0.913	1.184	0.997
YM3-KF	0.322	**0.723**	**0.934**	**0.998**	**0.912**	**0.988**	1.064	1.042	1.022	0.947	0.857
YM4-KF	**0.321**	0.725	0.944	1.115	1.122	1.092	**0.966**	**0.9043**	**0.921**	**0.922**	**0.822**

Table 7: Out-of-sample forecasting results US

	RMSE									Mean RMSE	MAE
Maturity	3m	1	2	5	10	15	20	25	30		
Panel A: 1-month horizon											
RW	**0.037**	0.408	0.212	0.084	0.319	0.436	0.223	0.202	**0.026**	**0.198**	**0.198**
YO-OLS	0.373	0.131	0.131	**0.026**	0.238	0.365	0.320	0.221	0.116	0.222	0.222
YO-KF	0.127	**0.049**	0.273	0.358	**0.164**	**0.043**	**0.086**	**0.193**	0.295	0.201	0.201
YM3-KF	0.371	0.129	**0.129**	0.027	0.235	0.360	0.315	0.216	0.110	0.223	0.223
YM4-KF	0.372	0.130	0.130	0.027	0.236	0.362	0.317	0.218	0.112	0.225	0.225
Panel B: 1-year horizon											
RW	0.948	0.806	0.560	0.655	0.976	0.525	1.112	0.439	0.909	0.928	0.792
YO-OLS	1.709	1.831	**0.792**	0.849	**0.490**	**0.488**	0.537	0.587	0.624	0.816	0.697
YO-KF	1.427	0.915	1.171	**0.726**	0.521	0.497	**0.540**	0.587	0.623	0.732	0.572
YM3-KF	**0.486**	0.530	0.865	0.770	0.563	0.523	0.624	0.787	0.928	0.678	0.548
YM4-KF	0.492	**0.517**	0.842	0.763	0.565	0.519	0.611	**0.768**	0.907	**0.641**	**0.538**
Panel C: 5-year horizon											
RW	1.822	2.389	2.084	0.954	1.612	1.187	1.441	1.057	1.763	1.446	1.124
YO-OLS	1.665	1.770	1.677	1.370	1.193	1.120	1.057	1.021	1.011	1.196	0.926
YO-KF	1.654	1.762	1.672	1.368	1.193	1.121	1.058	1.022	1.011	1.195	0.927
YM3-KF	0.317	0.472	0.623	0.909	1.124	1.206	1.249	1.185	1.117	0.982	0.912
YM4-KF	**0.259**	**0.440**	**0.572**	**0.778**	**0.957**	**1.018**	**1.037**	**0.954**	**0.975**	**0.871**	**0.681**

Table 8: Out-of-sample forecasting results EU

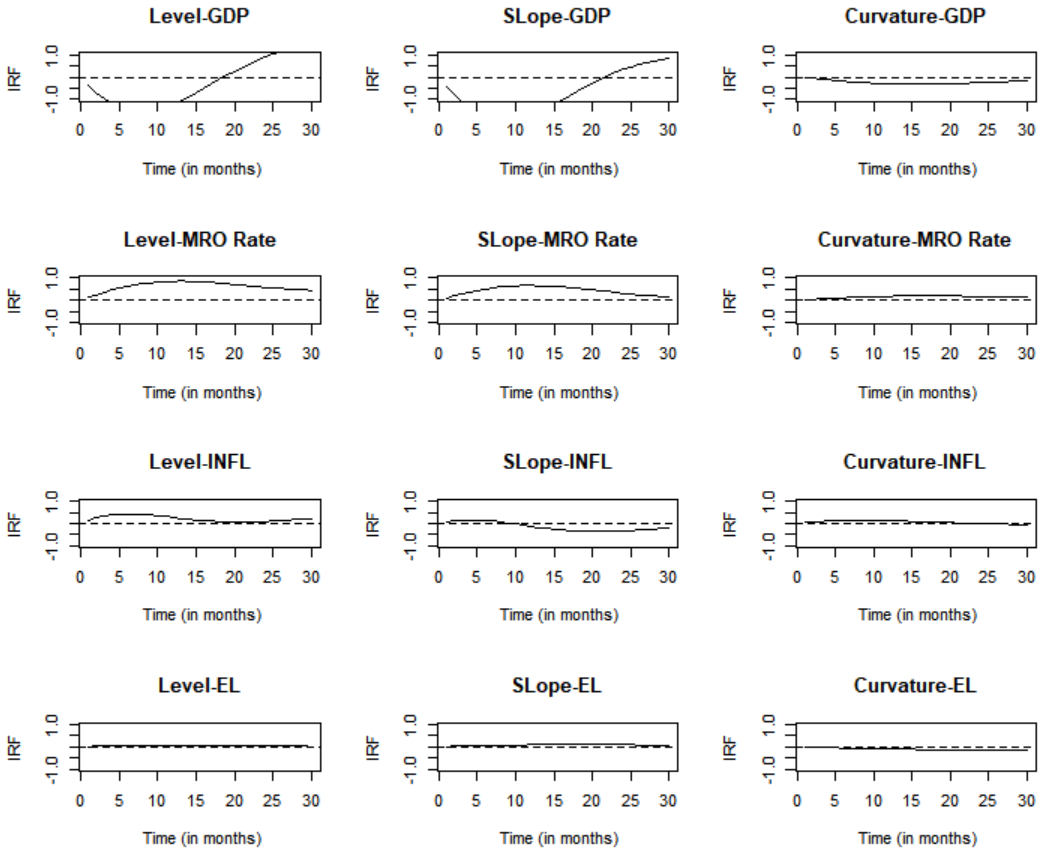


Figure 11: Impulse Response Function of the EA YM4-KF Model, depicting the dynamic impact of macro variables as the impulse on the model, and the subsequent response of the latent factors.

8 Stationarity and forecast sensitivity

This section assesses whether the model meets its underlying assumptions through examinations for stationarity and residual diagnostics. If evidence indicates non-stationarity, corrective transformations are applied to the series. Following this, an analysis of the sensitivity in forecasting performance for the nonstationary models is conducted and subsequently compared.

8.1 Stationarity conditions

Stationarity refers to the consistency of statistical properties in a process that generates a time series, indicating that these properties remain unchanged over time. In the context of multivariate modeling, a stable process is characterized by constant means, variances, and covariance over time, which implies stationarity. A VAR model is defined as stable when all eigenvalues of the parameters of autoregressive matrices are less than one in absolute terms. Furthermore, it's presupposed that each variable within the VAR framework adheres to a stationary process (Korstanje, 2021). The stationarity of a VAR model is fundamental for ensuring the reliability of its estimations, inferences, and future predictions, as it embodies resilience to shocks and a tendency to revert to a long-term mean or trend. Without stationarity, shocks could yield persistent or permanent impacts on the system. It's implausible, for example, that a disturbance in the policy rate from 1980 would continuously influence today's yield curve. Stationarity of the individual processes is a requirement for the stability of a VAR model. However, it does not guarantee the stability of the VAR model overall. To enforce the stability of the VAR model, we follow the work of Koopman et al. (2010) and apply the methodology of Ansley and Kohn (1986). This methodology enforces the stability of the VAR model by proposing a transformation of the transition matrix of a VAR model and its associated variance-covariance matrix. The procedure depends on finding a one-to-one mapping of the autocorrelation matrices of the matrices. For a comprehensive guide of this procedure, see the work of Ansley and Kohn (1986).

The foundational assumption made by Ansley and Kohn (1986) is that the individual series come from a stationary autoregressive process of order p. process. Accordingly, each macro variable in the model is visually assessed, followed by applying different statistical tests to check for stationarity. In instances of non-stationarity, differencing is employed, followed by subsequent reassessment for stationarity. The same is done for the assessment of the latent factors. The statistical tests employed to assess stationarity are the Augmented Dickey-Fuller (ADF) test, the Phillips-Perron (PP) test, and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Combining multiple statistical tests leverages the distinct null hypotheses and sensitivities of each test, thereby providing a comprehensive understanding of the stationarity characteristics of each series. In both the ADF and PP tests, the null hypothesis assumes that the process has a unit root, indicative of a non-stationary process, in contrast to the KPSS test. This test examines the stationary of a process around a deterministic trend, where the null hypothesis assumes that the process is trend-stationary.

Upon visual examination of the processes of the EA estimated latent factors in figure 10, it appears that the slope and curvature factors might be characterized as stationary, while the level factor exhibits a potential trend. The variable GDP, displayed in figure 12, potentially exhibits seasonal fluctuations, albeit not markedly prominent. The variables $CU$and FFR

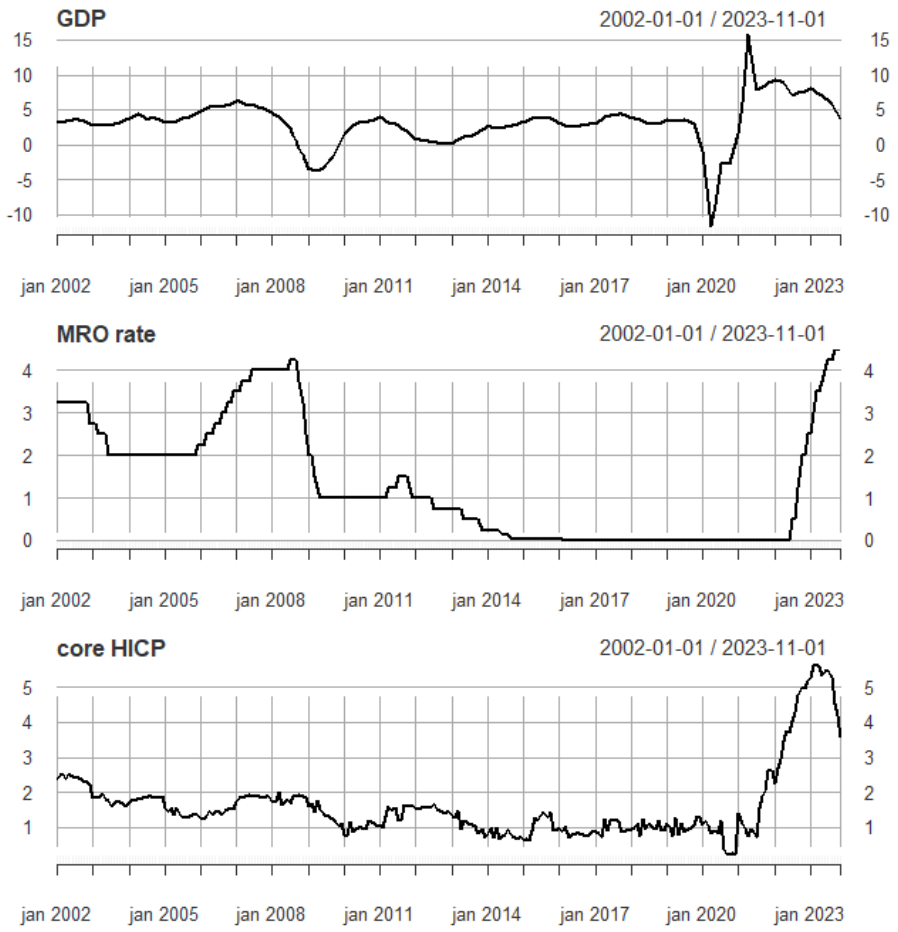


Figure 12: Time series of the key macroeconomic indicators of the EA

exhibit potential seasonal fluctuations, albeit not markedly prominent. The statistical tests in table 9 confirm some of the initial observations. True or False corresponds to whether the test rejects H0 or not at the 95% level. Failing to reject H0 for the ADF and PP test indicates evidence for non-stationarity. Failing to reject H0 in the case of the KPSS test indicates evidence for trend stationarity for the respective processes. Confirming these initial observations, statistical tests indicate evidence for trend-stationarity within the level factor, suggesting that the level factor is stationary after correcting for the trend. However, the results for the EA curvature factor are inconclusive since the ADF test fails to reject the null hypothesis, in contrast to the PP test. For all macro variables in both regions, the null hypothesis is not rejected in both the ADF test and the PP test, suggesting evidence of non-stationarity in all processes. Excluding the EL variable, there is also no evidence for trend stationarity. Combining all tests suggests that all processes are non-stationary or trend-stationary. By applying the first difference transformation to each series, each series is transformed into a stationary series. After applying the first-difference transformation to the data, there is conformity across the different statistical tests that the transformed processes are stationary. The graphs of the transformed level, slope, and curvature factors, as visually represented in figure 13, visually align with characteristics indicative of stationary behavior. The results of the statistical tests of both regions are displayed in table 10.

Panel A	Level	Slope	Curvature	CU	FFR	PCE	BSP
ADF	False	False	True	False	False	False	False
PP	False	False	True	False	False	False	False
KPSS	False	True	True	True	True	True	True
Panel B	Level	Slope	Curvature	GDP	MRO	HICP	EL
ADF	False	False	False	False	False	False	False
PP	False	False	True	False	False	False	False
KPSS	False	True	True	True	True	True	False

Table 9: Test for stationarity in level data. Results for the US are displayed in panel A. Results for the EA are displayed in panel B

Panel A	ΔLevel	ΔSlope	ΔCurvature	ΔCU	ΔFFR	ΔPCE	ΔBSP
ADF	True	True	True	True	True	True	True
PP	True	True	True	True	True	True	True
KPSS	True	True	True	True	True	True	True
Panel B	ΔLevel	ΔSlope	ΔCurvature	ΔGDP	ΔMRO	ΔHICP	ΔEL
ADF	True	True	True	True	True	False	True
PP	True	True	True	True	True	True	True
KPSS	True	True	True	True	True	True	True

Table 10: Test for stationarity in differenced data. Results for the US are displayed in panel A. Results for the EA are displayed in panel B

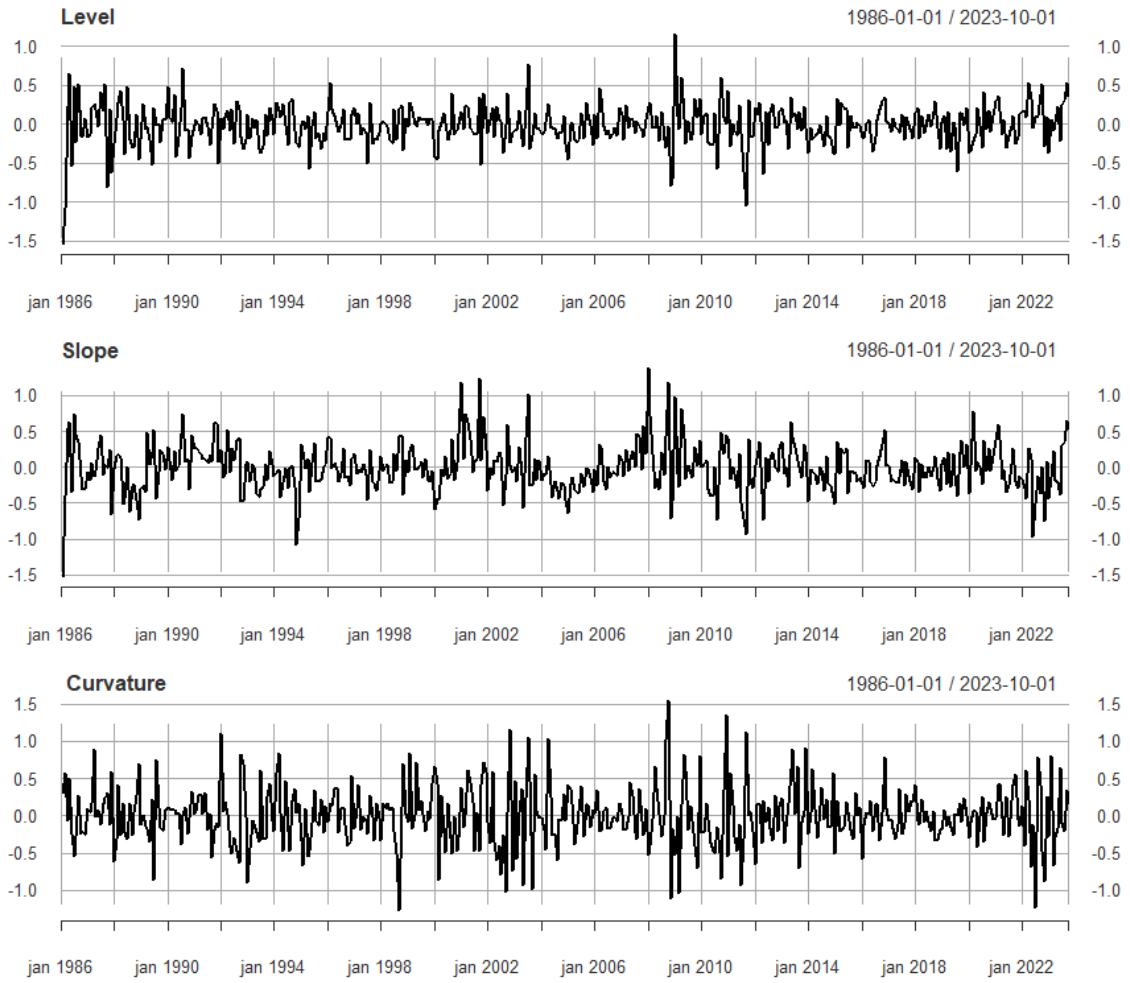


Figure 13: First order difference of the level, slope, and curvature

8.2 First-order difference DNS in state-space form

Taking the first difference transforms each individual process into a stationary process. Applying this transformation in the 2-step approach is straightforward once the series of each latent factor is obtained. However, in the 1-step approach, where the parameters and the latent factors are obtained simultaneously, it is less obvious how to facilitate this. However, taking the first difference in the interest rates facilitates this. Taking the first difference of the yields, transform the state-space model into:

$$\begin{pmatrix} \Delta y_t(\tau_1) \\ \Delta y_t(\tau_2) \\ \vdots \\ \Delta y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} \\ 1 & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} - e^{-\tau_N\lambda} \end{pmatrix} \begin{pmatrix} \Delta L_t \\ \Delta S_t \\ \Delta C_t \end{pmatrix} + \begin{pmatrix} \tilde{\epsilon}_t(\tau_1) \\ \tilde{\epsilon}_t(\tau_2) \\ \vdots \\ \tilde{\epsilon}_t(\tau_N) \end{pmatrix}, \tag{8.1}$$

$$\begin{pmatrix} \Delta L_t \\ \Delta S_t \\ \Delta C_t \end{pmatrix} = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \tilde{a}_{13} \\ \tilde{a}_{21} & \tilde{a}_{22} & \tilde{a}_{23} \\ \tilde{a}_{31} & \tilde{a}_{32} & \tilde{a}_{33} \end{pmatrix} \begin{pmatrix} \Delta L_{t-1} \\ \Delta S_{t-1} \\ \Delta C_{t-1} \end{pmatrix} + \begin{pmatrix} \tilde{\eta}_{1,t} \\ \tilde{\eta}_{2,t} \\ \tilde{\eta}_{3,t} \end{pmatrix}. \tag{8.2}$$

Since the macro variables are treated as exogenous variables in the models, extending the above model with differenced macro variables is a straightforward extension and analog to 4.1 and 4.2. Again, orthogonality between the white noise errors is needed to find optimality in the Kalman estimation procedure. Also, diagonal H and unrestricted Q matrix are assumed.

8.3 Out-of-sample performance

In this subsection, each model's out-of-sample performance forecasting performance wiht stationary data is assessed for each of the three forecast horizons. Again, the random walk model is used as the benchmark. Subsequently, the error terms of the best-performing models in levels and first differences are tested to determine whether they adhere to the assumption of being white noise by assessing their autocorrelation and stationary. Autocorrelation and stationarity of the error terms are being assessed by the LB and ADF tests.

8.3.1 1-month horizon

Panel A of table 11 and 12 display the 1-month forecast results. The results reveal a consistent trend across regions, where all models demonstrate superior performance when estimated on differenced data compared to equivalent models estimated in levels. Notably, the models for the US exhibit a slight advantage over the random walk, further highlighting the effectiveness of using stationary data. In the models applied to the data of the EA, it still fails to outperform the random walk. Notably, the YO-KF-diff consistently emerges as the top-performing model in both regions, surpassing models incorporating macro variables and the YO-OLS-diff model. The biggest reduction in RMSE is in the models incorporating macro variables. This again indicates that noise is reduced for this horizon.

8.3.2 1-year horizon

The results in panel B reflect the forecast performance of the proposed models of the 1-year forecasting horizon. Each proposed model is superior to the random walk model in terms of mean RMSE and MAE. The absolute difference between the models' forecast errors and the random walk's forecast errors also increases for longer forecast horizons, implying that the forecasting accuracy of the models improves over time. Notably, the YM4-KF-diff model has the lowest out-of-sample error, demonstrating the highest forecasting accuracy, with a mean RMSE of 0.550 and 0.682 in the US and EA, respectively. The order of model performance remains consistent with models estimated in levels, suggesting that the inclusion of specific macro variables and estimation within the state space framework enhances forecast accuracy, even when the data is differenced. A particular highlight is its good fit to the EA's YM4-KF-diff to the front end of the swap curve. a stark contrast emerges when evaluating the performance of the 1-month horizon models in comparison to their equivalent counterparts that do not account for stationarity. In this context, models estimated on level data consistently outperform their differenced data counterparts. The most notable distinctions between equivalent models are observed prominently in instances where macro variables are not incorporated.

8.3.3 5-year horizon

In the comprehensive evaluation of forecast performance for the 5-year horizon, displayed in panel C of both tables, the observations are comparable with those of the 5-year horizon models estimated in levels. Again, the random walk model is outperformed by all other models, underlining the consistent superiority of the proposed forecasting methodologies. Furthermore, the widening differences between the random walk model and alternative models indicate a notable enhancement in forecasting accuracy for longer forecast horizons compared to the random walk benchmark. Also, the YO model estimated based on the Kalman filter outperforms the YO model estimated by OLS, further strengthening the case for using the state-space framework in this context. The YM4-KF-diff model stands out as the best-performing model in both regions, demonstrating superiority across various maturities. This model's dominance extends to almost all maturities, underscoring its robust forecasting capabilities for longer forecast horizons. Enforcing the results of the 1-year horizon forecast errors, the models estimated on level data consistently outperform their counterparts estimated on differenced data for the 5-year forecast horizons. YM4-KF is superior to the YM4-KF-diff model with lower mean RMSE of 0.051 and 0.078 basis points for the US and EA, respectively.

	RMSE									Mean RMSE	MAE
Maturity	3m	1	2	5	10	15	20	25	30		
Panel A: 1-month horizon											
RW	0.244	**0.072**	**0.004**	0.122	0.242	0.298	0.279	0.198	**0.076**	0.199	0.199
YO-OLS-diff	**0.106**	0.194	0.189	**0.032**	**0.034**	**0.121**	**0.212**	0.208	0.127	0.193	0.193
YO-KF-diff	0.094	0.133	0.156	0.153	0.112	0.184	0.216	**0.176**	0.159	**0.182**	**0.182**
YM3-KF-diff	0.092	0.090	0.111	0.129	0.134	0.173	0.251	0.231	0.216	0.201	0.201
YM4-KF-diff	0.091	0.087	0.142	0.130	0.135	0.173	0.249	0.229	0.210	0.198	0.198
Panel B: 1-year horizon											
RW	**0.111**	1.103	1.371	**0.421**	0.745	0.701	1.07	**0.494**	0.99	0.876	0.757
YO-OLS-diff	0.328	0.431	0.371	0.719	0.921	0.928	0.910	0.910	0.937	0.818	0.752
YO-KF-diff	0.330	**0.437**	0.362	0.664	0.839	0.834	0.809	0.803	0.825	0.740	0.668
YM3-KF-diff	0.841	0.745	0.442	0.492	0.635	**0.586**	**0.532**	0.500	**0.498**	0.563	0.508
YM4-KF-diff	0.531	0.452	**0.306**	0.463	**0.613**	0.606	0.579	0.568	0.585	**0.550**	**0.493**
Panel C: 5-year horizon											
RW	1.694	1.833	1.787	1.576	3.465	1.068	1.511	1.551	3.126	1.939	1.645
YO-OLS-diff	1.993	1.822	1.693	1.693	1.794	1.769	1.716	1.664	1.620	1.735	1.441
YO-KF-diff	1.866	1.768	1.587	1.324	1.318	1.332	1.313	1.282	1.249	1.370	1.108
YM3-KF-diff	0.273	**0.568**	**0.660**	**0.810**	**1.043**	1.136	1.155	1.145	1.127	0.992	0.794
YM4-KF-diff	**0.272**	0.559	0.663	0.836	1.044	**1.112**	**1.115**	**1.095**	**1.071**	**0.973**	**0.786**

Table 11: Out-of-sample forecasting results for stationary models US

	RMSE									Mean RMSE	MAE
Maturity	3m	1	2	5	10	15	20	25	30		
Panel A: 1-month horizon											
RW	**0.037**	**0.408**	0.212	**0.084**	0.319	0.436	0.223	0.202	**0.026**	**0.198**	**0.198**
YO-OLS-diff	0.020	0.100	0.038	0.150	0.254	0.298	0.284	0.233	**0.164**	0.210	0.210
YO-KF-diff	0.292	0.418	0.374	0.245	**0.194**	**0.156**	**0.148**	**0.163**	0.190	0.205	0.205
YM3-KF-diff	0.030	0.092	**0.031**	0.156	0.260	0.304	0.291	0.240	0.171	0.215	0.215
YM4-KF-diff	0.044	0.046	0.041	0.266	0.302	0.358	0.342	0.277	0.185	0.234	0.234
Panel B: 1-year horizon											
RW	0.948	0.806	0.560	0.655	0.976	0.525	1.112	0.439	0.909	0.928	0.792
YO-OLS-diff	2.310	2.405	2.062	1.338	0.996	0.873	0.670	0.466	0.311	0.920	0.928
YO-KF-diff	2.020	2.163	1.952	1.359	0.964	0.790	**0.579**	**0.387**	**0.250**	0.889	0.858
YM3-KF-diff	0.270	**0.462**	**0.628**	**0.683**	**0.743**	0.811	0.930	1.082	1.224	0.845	0.657
YM4-KF-diff	**0.268**	0.498	0.693	0.757	0.744	**0.716**	0.658	0.656	0.699	**0.682**	**0.552**
Panel C: 5-year horizon											
RW	1.822	2.389	2.084	0.954	1.612	1.187	1.441	1.057	1.763	1.446	1.124
YO-OLS-diff	1.488	1.615	1.530	1.258	1.261	1.240	1.228	1.259	1.311	1.288	1.116
YO-KF-diff	1.560	1.687	1.617	1.323	1.219	1.198	1.187	1.218	1.274	1.275	1.097
YM3-KF-diff	0.457	**0.653**	**0.816**	**0.827**	1.031	1.081	1.146	1.243	1.344	1.052	0.943
YM4-KF-diff	**0.421**	0.861	1.235	1.150	**0.966**	**0.884**	**0.786**	**0.719**	**0.691**	**0.951**	**0.832**

Table 12: Out-of-sample forecasting results for stationary models EU

8.3.4 Assessment white noise

Applying the first-order difference transformation to the data reveals that, in most instances, there is evidence to classify the error terms as white noise by combining the results of the two tests. In all cases, the ADF test rejects the null hypothesis, suggesting no presence of unit root. In the case of the LB test, the test fails to reject H0 in almost all cases, indicating no autocorrelation of the error terms.

The application of a first-order difference transformation to the data provides evidence supporting the classification of error terms as white noise, as inferred from the combined results of the two tests. Across all instances, the ADF test rejects the null hypothesis, indicating the absence of a unit root. Simultaneously, the LB test fails to reject the null hypothesis in almost all cases, suggesting there is no autocorrelation in the error terms.

	YM4-KF		YM4-KF-diff	
Error	LB	ADF	LB	ADF
$\epsilon_t(\tau_1)$	True	False[a]	False	True
⋮	⋮	⋮	⋮	⋮
$\epsilon_t(\tau_N)$	True	False	False	True
η_1	False	True	False	True
η_2	True	False	False	True
η_3	True	False	False	True
η_4	True	False	False	True
η_5	True	False	True	True
η_6	True	True	False	True
η_7	True	True	False	True

(a) US, a: H0 is rejected for maturities t = 15,26,27,29

	YM4-KF		YM4-KF-diff	
Error	LB	ADF	LB	ADF
$\epsilon_t(\tau_1)$	True	False[a]	False[b]	True
⋮	⋮	⋮	⋮	⋮
$\epsilon_t(\tau_N)$	True	False	False	True
η_1	False	True	False	True
η_2	True	True	False	True
η_3	False	True	False	True
η_4	True	True	False	True
η_5	True	True	True	True
η_6	True	True	False	True
η_7	True	True	False	True

(b) EU, a: H0 is rejected for error terms corresponding to maturities t = 11,12,13,18 (implying for stationarity) ; b: H0 is rejected for error terms with maturities t = 26,27,29 are rejected(implying evidence for auto correlation)

Figure 14: Caption for the entire figure

9 Scenario analysis

This section focuses on examining the trajectory of interest rate curves across various horizons in both regions. The analysis is performed using the best-performing models identified for each specific region. Three distinct macroeconomic scenarios are considered. The first scenario illustrates the interest curve's trajectory based on the macroeconomic projections provided by the respective central banks. Following this, two contrasting extremes on the economic spectrum are included: a recession and an economic expansion scenario. Additionally, the chapter presents the trajectory of both the 10-year and short-term interest rates under these scenarios.

9.1 Scenarios

The generation of the different scenarios for each region follows a similar approach. The scenario based on projections of the respective central banks is considered the main scenario. For the three initial macro variables, the ECB provides annual projections for 2023 through 2026 and a long-term projection. The Fed does not directly provide yearly projections for CU, but it does for the FFR, PCE, and UR. Projections for CU are obtained by establishing and utilizing the relationship between CU and UR. The relationship is established by regressing UR onto CU. The resulting regression model is given by:

$$CU_t = 85.649 - 1.376\, UR_t. \tag{9.1}$$

Determining the fourth macroeconomic variable in the European and American markets is contingent upon forward guidance provided by their respective central banks. For instance, as part of the European QT policy, the ECB has announced a predefined trajectory for diminishing assets acquired under the PEPP. The reduction involves a monthly decrease of 7.5 billion euros in assets purchased under this program, with the intention to cease reinvestments under the PEPP at the conclusion of 2024.

The formulation of the other two scenarios involves examining extreme values (90th and 10th percentiles) within the data and considering the interactions between economic variables based on established economic theory. During periods of economic expansion, expectations align with heightened economic activity and increased inflation. To counter potential overheating, strict monetary policies are maintained. This is reflected in the EA's elevated GDP, policy rates, and inflation assumptions, accompanied by a parallel expectation for the US. The latter includes a low unemployment rate. In addition to standard monetary policy, anticipations encompass the implementation of a strict QT policy, reducing the central banks' balance sheets and subsequently reducing the available liquidity in the systems.

Conversely, economic recessions are often characterized by low to negative economic growth, low inflation, and loose monetary policies. In the EA, this translates to diminished GDP, low inflation, and a reduced policy rate. Simultaneously, excess liquidity in the system is anticipated to be higher. Parallel characteristics are assumed for the US, featuring a high unemployment rate combined with low inflation in accordance with the Philips curve. Tables 13 and 14 present the projections of the macro variables under each scenario for the US and the EA, respectively.

Scenario	Macro variables	2024	2025	2026	2027	2028
Main (Fed)	CU	80.4	80.0	80.0	80.0	80.1
	(UR)	(3.8)	(4.1)	(4.1)	(4.1)	(4)
	FFR	5.4	4.6	3.6	2.9	2.5
	INFL	3.7	2.4	2.2	2	2
	BSP	17	13.5	12.9	12.5	12.2
Economic Expansion	CU	81.5	81.5	81.5	81.5	81.5
	(UR)	(3)	(3)	(3)	(3)	(3)
	FFR	6	7	6	5	4
	INFL	5	5	4	4	3
	BSP	16	12	9	8	7.5
Recession	CU	78.8	76.4	74.6	74.6	78.8
	(UR)	(5)	(8)	(8)	(8)	(5)
	FFR	2	1	1	1	1
	INFL	1	0.5	0.5	0.5	1
	BSP	18	22	26	28	30

Table 13: Macroeconomic projections in the US under each scenario, reflecting values as of December 1st

Scenario	Macro variables	2023	2024	2025	2026	2028
Main (ECB)	GDP	3.1	0.8	1.5	1.5	3
	MRO	4.5	3.6	2.8	2.7	1.5
	INFL	3.2	2.7	2.1	1.9	2
	EL	1	0.75	0.6	0.5	0.5
Economic Expansion	GDP	5	6	6	5	3.5
	MRO	5	5	5	4	2
	INFL	5	5	5	4	3
	EL	-0.7	-0.7	-0.7	-0.7	0
Recession	GDP	2	0	0	0	1.5
	MRO	4	0	0	0	1
	INFL	2.5	0	0	0	1
	EL	2	2	2	2	1

Table 14: Macroeconomic projections in the EA under each scenario, reflecting values as of December 1st

9.2 Forecasts

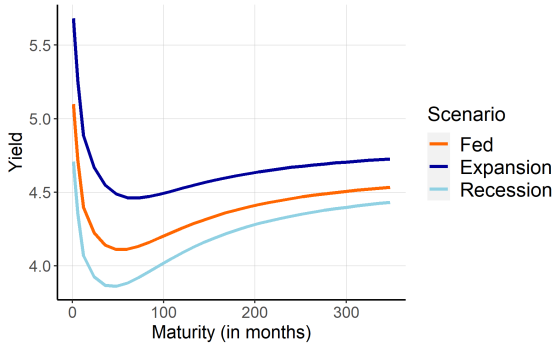
The US and EA forecasts under each scenario and for two forecast horizons are presented in figure 15 and figure 16, respectively. Both graphs exhibit comparable characteristics, showcasing the models' ability to fit various term structure shapes. The model predicts a persistent inversion of the yield curve in both regions in the next period, gradually transitioning to a traditionally upward-sloping term structure. The latter is caused by a decrease in the front end and an increase in the back end of the term structure in each specific region.

Due to their practical economic significance, particular attention is dedicated to the short-rate and 10-year rates. These rates serve as crucial indicators. The 10-year rate acts as a barometer for other key rates, such as mortgage rates, and functions as a marker for investor confidence. The short-term rate serves as a benchmark for the policy rate, exhibiting high responsiveness to changes in the policy rate with concurrent directional movements. Due to the nonsmooth nature of the policy rate and the close relationship with the short-term rate, a significant contrast in smoothness is observed between the forecasted trajectories of the short-rate and the 10-year rate, as depicted in Figures 15c, 15d, 16c, 16d. Another vital indicator of economic health is reflected in whether the term structure is inverted. This study defines inversion of the term structure as the condition where the spread between the 10-year and 3-month rates in the US, and the 10-year and 6-month rates in the EA, is negative.

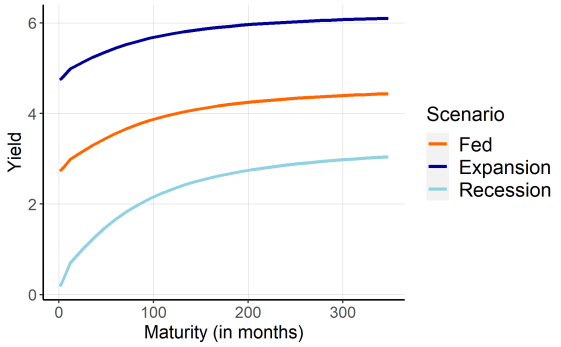
If the economy evolves in line with the projections of the Fed and the ECB, similar patterns emerge for the respective rates in the US and the EA. In the US, the Treasury yield is expected to converge to the average yield curve by 2028, and the yield curve is expected to slope upwards by July 2024. The 10-year rates are foreseen to peak around 5% in 2025, followed by a gradual decline towards the 3-4% range in subsequent years. In the EA, projections suggest that by 2028, the swap curve will be positioned above the average curve, featuring a less steep slope. The inversion in the swap curve is expected to reverse by the end of 2024, with the 10-year swap rates reaching a peak of 3.7% in 2025, followed by a gradual descent to 3% in the following years.

In the event of an economic expansion scenario, the outlook diverges further. In the US, under this expansion scenario, the curve is expected to normalize by early 2025, with the 10-year rate obtaining a maximum of 6.2% in October 2025. The term structure in the EA is expected to normalize by mid-2025, with the 10-year swap rate peaking at around 4.8% in 2026. In both regions, the term structure is positioned above the term structure of the main scenario.

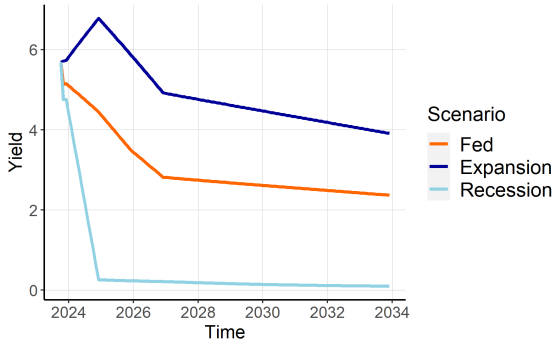
In the context of an economic recession scenario, the inversion of the Treasury curve is expected to revert around mid-2024. The 10-year Treasury rate is expected to decline immediately, moving towards 1% in subsequent years. In the EA, the swap curve is expected to be upward-sloping by March 2024, with the 10-year rate declining towards 1%, whereas it will show a little increase from 2030. Both term structures lie below the term structure of the main scenario



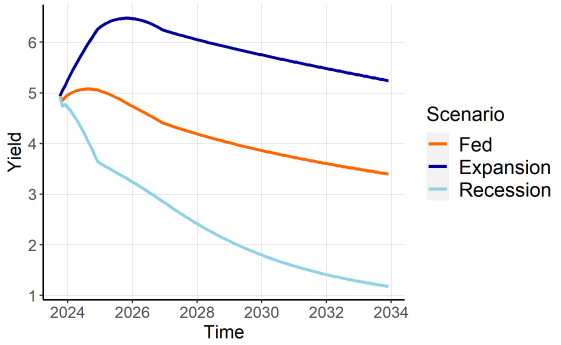
(a) 01-12-2023



(b) 01-12-2028

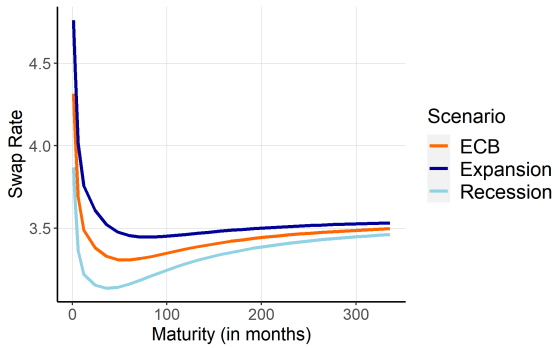


(c) 3-month yield

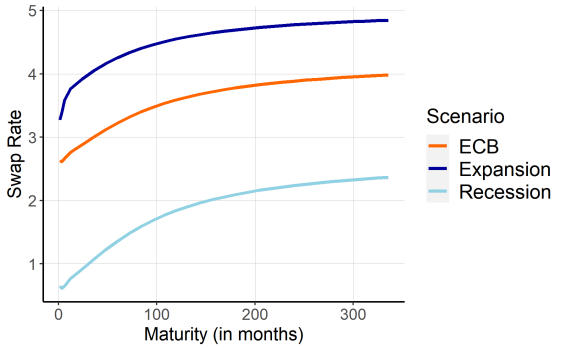


(d) 10-years yield

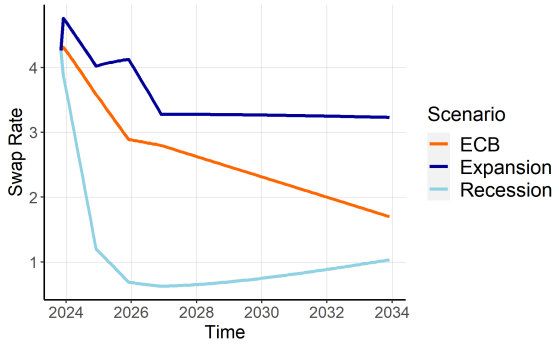
Figure 15: YM4-KF Yield Curve US under three different scenarios with data up to 2023-08-01 and Term Structure of Interest Rates for 2023 and 2033



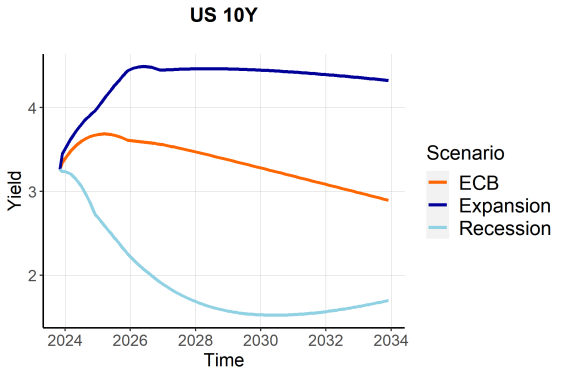
(a) 01-12-2023



(b) 01-12-2028



(c) 6-month swap rate



(d) 10-years swap rate

Figure 16: YM4-KF Yield Curve US under three different scenarios with data up to 2023-08-01 and Term Structure of Interest Rates for 2023 and 2033

10 Discussion

10.1 Concluding summary

In this study, I investigated the out-of-sample forecasting performance of two extensions of the dynamic Nelson-Siegel model (DNS) proposed by Diebold et al. (2006) and applied it to Treasury yields in the United States and 6-month Euribor swap rates in the EA. Each extension's forecast accuracy is tested over different horizons and benchmarked against other versions. The first extension incorporates a measure for the balance sheet of the respective central bank as the fourth macroeconomic variable into the DNS model. Although the balance sheet policy is an unconventional monetary policy, it has been used prominently since 2008 and significantly impacts the term structure (Dröes et al., 2017; Krishnamurthy & Vissing-Jorgensen, 2011) . Once the relationship between the term structure and the macroeconomic variables is established, it allows for forecasting the trajectory of the term structure under different economic scenarios. The second extension focuses on evaluating the forecasting performance and sensitivity of the model when addressing non-stationarity in latent factors and macroeconomic variables. This extension could be relevant since the original DNS model relies on the assumption that the latent factors and macro variables follow a stationary autoregressive process; however, in the existing literature, this issue has largely been ignored.

The selection and construction of the macroeconomic variables used in the US and EA differ slightly. The reasons are that the Fed formulates its monetary policy primarily based on domestic economic conditions, while the ECB exercises influence over the monetary policy of a diverse group of member countries, each characterized by unique economic conditions and equipped with its own national central bank, thereby introducing additional complexities. Furthermore, the implementation of QE and QT policies differs regarding the targeted assets and implementation protocols. Initially, various potential measures for the balance sheet were considered for each region and evaluated based on RMSE. Given the structural disparities between the US and the EA, two distinct yet interconnected measures for the central balance sheet are employed for the US and the EA. The balance sheet of the Fed is measured as the ratio of the value of Treasuries and MBS the bank possesses to the total value of Treasuries and MBS outstanding in the market. In the EA, excess liquidity serves as a measure for the balance sheet of the ECB. Additionally, a six-month lead for both measures is incorporated to account for the effects of the initial announcements of their policies.

Subsequently, the out-of-sample forecast performance of the models is compared across different forecasting horizons (1 month, 1 year, and 5 years). The comparison includes benchmarking against the random walk model. Forecast accuracy is assessed through the RMSE for each maturity, and the overall fit is determined by averaging the RMSE values for each maturity. Consistent performance is observed across all models when applied to datasets from the US and the EA. The findings suggest comparable effectiveness of the models in both regions. For the 1-month forecast horizon, none of the models demonstrated superior performance compared to the random walk model. Notably, the YO-KF model emerged as the second-best performer at this horizon, outperforming models incorporating macroeconomic variables. This contrasts with the results for the other two forecast horizons, where the YM4-KF model exhibited superior performance compared to all other models. The observed superiority of the YM4-KF model in longer forecast horizons indicates the significance of including the central bank's balance sheet as an additional variable. This inclusion is seen as providing valuable

information, particularly for extended forecasting periods. The random walk model demonstrates effectiveness in capturing short-term randomness, while models like DNS, particularly when augmented with macro variables, are designed to capture more complex relationships over longer horizons, as evidenced by their superior forecasting performances over longer horizons.

Next, the analysis shifts to examining the implicit and explicit assumptions inherent tp the state space model. The initial implicit assumption suggests that the state space model is built on a stable process, indicating that the latent factors exhibit stationarity. A second assumption relates to the error terms, assuming they follow white noise characteristics. However, it becomes clear that both assumptions are not strictly followed, as there is empirical evidence of non-stationarity in many variables, and the error terms do not show white noise characteristics. To address these issues, the first-order differencing is employed to the data as a corrective measure. The subsequent models outperform those models estimated on level data for the 1-month forecast horizon. However, for longer forecast horizons, the models estimated on level data demonstrate superiority over those estimated on differenced data. This implies that differencing nonstationary data effectively captures short-term dynamics by eliminating long-term trends and relationships. Nevertheless, the consequent loss of information makes models estimated on level data superior for longer forecast horizons.

Finally, scenario analysis is conducted using the YM4-KF model, chosen for its superior out-of-sample forecasting performance over extended horizons. The term structures in both the US and the EA exhibit a comparable trajectory, featuring an inverted shape in December 2023 and a transition towards an upward slope by 2028. The influence of monetary policy on the term structure is notably evident through the incorporation of the policy rate and a measure of the central bank's balance sheet. Projections of the term structure until 2028 are explored across three distinct economic scenarios. The main scenario relies on macroeconomic variable projections provided by the respective central banks. The term structure projections align with traditional economic expectations, wherein economic expansion leads to increased consumer spending, inflation, and a subsequent rise in the term structure. Conversely, a recession anticipates a decrease in the respective term structure. The two alternative scenarios can be interpreted as the upper and lower bounds of potential interest rate movements. In addition, insights can be obtained into when the term structure will revert to an upward-sloping curve and how far interest rates could rise. If the economy evolves in line with the projections of the respective central bank, the term structures are expected to revert to a standard upward-sloping graph in mid-2024 for the US and at the end of 2024 for the EA. Observable is the trend that the term structure in both regions tends to shift back towards an upward-sloping graph earlier during periods of recession compared to times of expansion.

11 Limitations and future research

This section discusses the limitations that have emerged from the study and suggestions for future research. Some of these limitations can serve as potential areas for future research.

This model's first limitation arises from using exogenous macro variables to recursively forecast the VAR framework's latent factors. In this context, while the model effectively captures the influence of realized macroeconomic variables on the latent factors, it does not permit the endogenous evolution of these macro variables within the system. This simplification may overlook potential feedback loops and interactions between latent factors and macro variables, limiting the model's ability to account for dynamic relationships fully. In addition, this simplification might affect the accuracy of my forecasts, especially in situations where the real relationship involves these factors influencing each other.

A secondary limitation, which is simultaneously open for exploration in future research, stems from the methodology employed, which restricts the decay parameter from changing over time. Previous studies that applied the DNS model to different time periods found different values for lambda, reflecting the variability in decay behavior over time. To address this limitation, one could consider treating lambda as an additional dynamic factor, as demonstrated by Koopman et al. (2010), allowing lambda to change over time. Alternatively, a regime shift model could be introduced to account for variations in the decay parameter across different regimes.

The third limitation of the model pertains to the reliance on the quality of projections for macroeconomic variables used in scenario forecasts. The effectiveness of scenario forecasts is inherently tied to the accuracy and reliability of these projections. Given the complex nature of economic variables and the inherent uncertainties, any inaccuracies or limitations in the projections directly impact the robustness of the scenario forecasts. Therefore, the model's capacity to provide insightful scenario forecasts is contingent on the availability of precise and well-informed projections for the macroeconomic variables under consideration.

A final limitation, while also presenting an opportunity for future research exploration, relates to the use of differencing to address non-stationarity. Differencing can potentially result in the loss of information regarding the original levels of the series,particularly valuable for comprehending long-term trends. Recognizing this drawback and considering the enhanced forecasting performance for short horizons underscores the significance of exploring more robust modeling approaches that can correct for nonstationary without sacrificing long-term information.

In contrast to the limitations, the models and techniques used to estimate these models lay a good foundation for future exploration and improvements.

First, a good direction is to extend the model by incorporating an extra factor and decay parameter as proposed by Svensson (1994), which introduces more flexibility. This extension offers the potential for an improved in-sample fit, specifically addressing the second hump observed in the back end of the term structure. However, the introduction of an additional factor and an additional parameter raises the possibility of parameter instability. Investigating this further would be an interesting investigation to determine whether it leads to enhanced

out-of-sample forecasting or potentially indicates a case of overfitting the data.

Secondly, further generalization of the YM4-KF model by incorporating the concept of no-arbitrage as proposed by Christensen et al. (2009). This extension aims to create a more advanced framework for examining the interaction between macroeconomic factors and the term structure dynamics. It raises intriguing questions about the model's capacity to identify and explain market anomalies and whether this inclusion enhances the precision of pricing predictions.

References

Ansley, C. F., & Kohn, R. (1986). A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *Journal of Statistical Computation and Simulation*, *24*(2), 99–106.

Björk, T., & Christensen, B. J. (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance*, *9*(4), 323–348.

Black, F., & Scholes, M. (1972). The valuation of option contracts and a test of market efficiency. *The journal of finance*, *27*(2), 399–417.

Bolder, D. J., & Stréliski, D. (1999). Yield curve modelling at the bank of canada.

Braun, B. (2017). Central bank planning: Unconventional monetary policy and the price of bending the yield curve.

Brennan, M., & Schwartz, E. (1982). An equilibrium model of bond pricing and a test of market effciency. *Journal of Financial and Quantitative Analysis*, *17*, 301–329. https://doi.org/10.2307/2330832

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2009). An arbitrage-free generalized nelson–siegel term structure model.

Coroneo, L., Nyholm, K., & Vidova-Koleva, R. (2011). How arbitrage-free is the nelson–siegel model? *Journal of Empirical Finance*, *18*(3), 393–407.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*.

Dickey, D. A., & Fuller, W. (1981). *Distribution of likelihood ratio test statistics for nonstationary time series*. Citeseer.

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

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

Dröes, M., Lamoen, R., & Mattheussens, S. (2017). Quantitative easing and exuberance in government bond markets: Evidence from the ecb's expanded asset purchase program.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, *57*(1), 405–443.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, *6*(4), 379–406.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

Eser, F., Lemke, W., Nyholm, K., Radde, S., & Vladu, A. (2019). Tracing the impact of the ecb's asset purchase programme on the yield curve.

Gagnon, J., Raskin, M., Remache, J., & Sack, B. (2011). The financial market effects of the federal reserve's large-scale asset purchases. *International Journal of Central Banking*, *7*(1), 45–52.

Gagnon, J., Raskin, M., Remache, J., & Sack, B. P. (2011). Large-scale asset purchases by the federal reserve: Did they work? *Economic Policy Review*, *17*(1), 41.

Gambetti, L., & Musso, A. (2020). The effects of the ecb's expanded asset purchase programme. *European economic review*, *130*, 103573.

Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, 77–105.

Heston, S. L. (1993). A closed-form solution for options pricing with stochastic volatility. *The Review of Financial Studies*, *6*(2), 327–346.

Ho, T. S., & Lee, S.-B. (1986). Term structure movements and pricing interest rate contingent claims. *the Journal of Finance*, *41*(5), 1011–1029.

Hull, J., & White, A. (1990). Pricing interest-rate-derivative securities. *The review of financial studies*, *3*(4), 573–592.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

Koopman, S. J., Mallee, M. I., & Van der Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, *28*(3), 329–343.

Korstanje, J. (2021). *Advanced forecasting with python*. Springer.

Krishnamurthy, A., & Vissing-Jorgensen, A. (2011). *The effects of quantitative easing on interest rates: Channels and implications for policy* (tech. rep.). National Bureau of Economic Research.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, *54*(1-3), 159–178.

Loonen, R. (2022). *Estimating the effectiveness of u.s. quantitative easing programs and performing scenario analysis on the yield curve: A dynamic nelson siegel approach* [Unpublished master's thesis]. Tilburg University.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, 473–489.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *biometrika*, *75*(2), 335–346.

Pooter, M. D. (2007). *Examining the nelson-siegel class of term structure models* (tech. rep.). Tinbergen Institute Discussion Paper.

Schumacher, J. (2020). *Introduction to financial derivatives: Modeling, pricing and hedging*. Open Press TiU.

Smith, A. (2020). The european central bank's covered bond purchase programs i and ii (ecb gfc). *Journal of Financial Crises*, *2*(3), 382–404.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994.

Trichet, J.-C., & Papademos, L. (2009, May). Transcript of the questions asked and the answers given by jean-claude trichet, president of the ecb, and lucas papademos, vicepresident of the ecb. https://ypfsresourcelibrary.blob.core.windows.net/fcic/YPFS/2009-05-%2007%20Introductory%20statement%20with%20Q&A.pdf

van Dijk, M., Dubovik, A., et al. (2018). Effects of unconventional monetary policy on european corporate credit. *Discussion Papers*, (372).

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, *5*(2), 177–188.

A Plots

A.1 Average YC US

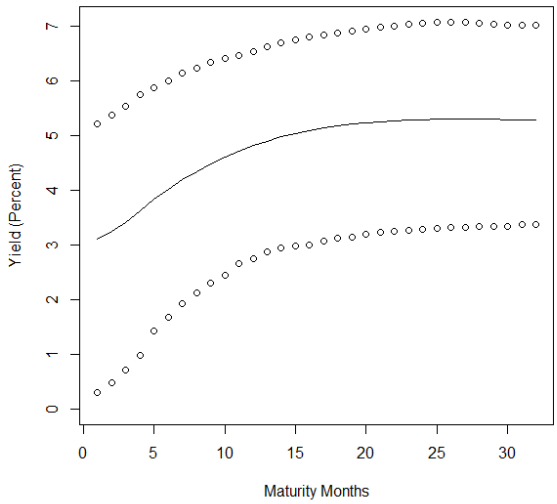


Figure 17: Behaviour of the loadings of the factors in the Nelson-Siegel model under different λ. Dotted line represents $\lambda = 0.035$. Solid line represents $\lambda = 0.08$

A.2 Latent factors US model vs emp

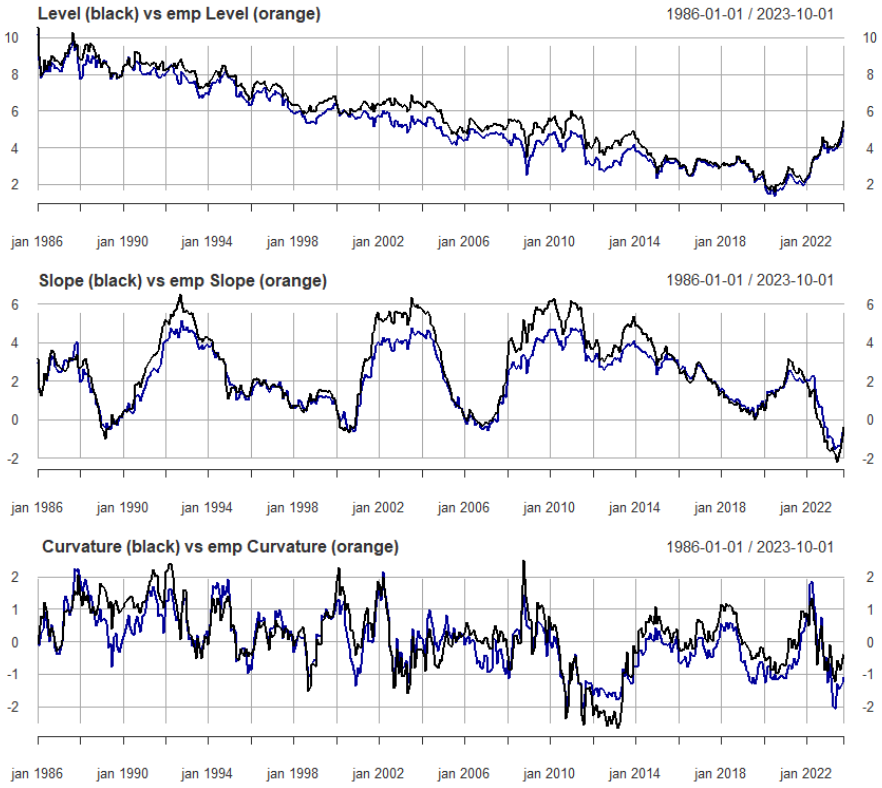


Figure 18: YM4-based level, slope and curvature and data-based level, slope and curvature.

A.3 Macroeconomic variables US

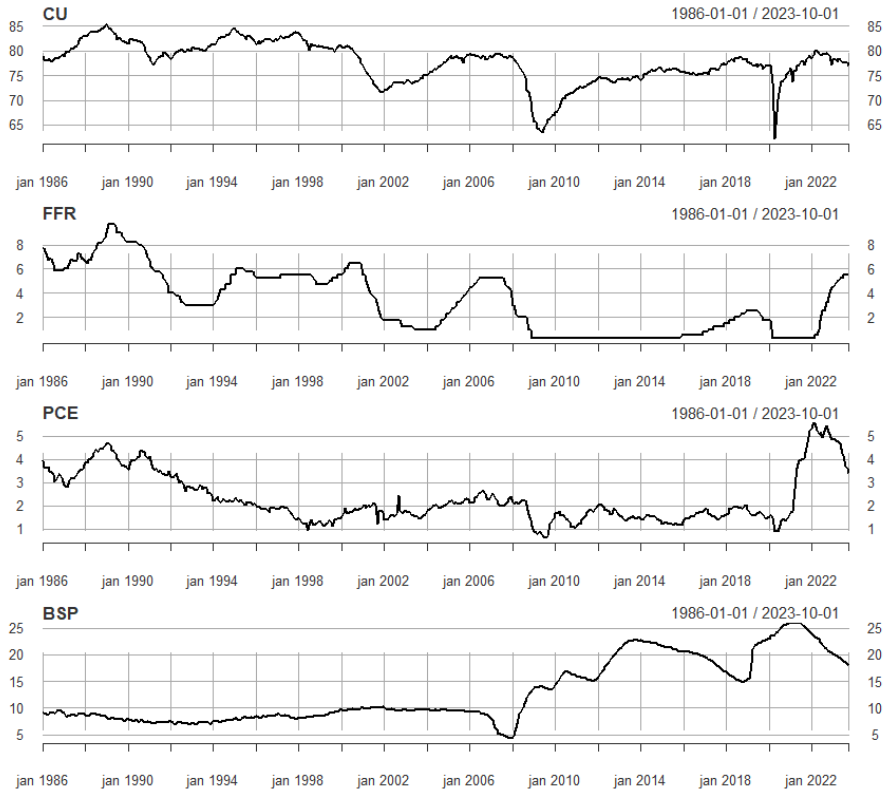


Figure 19: Time series macroeconomic variables US

A.4 IRF factors to macro variables

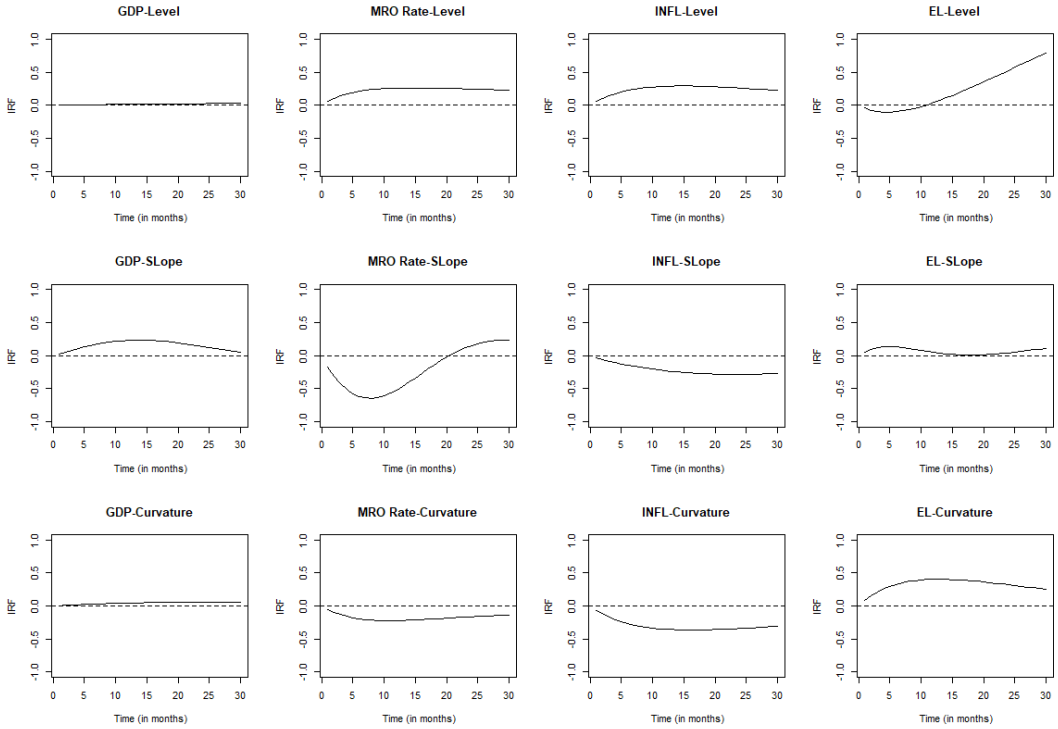


Figure 20: Impulse Response Function of the EA YM4-KF Model, depicting the dynamic impact of latent factors as the impulse on the model, and the subsequent response of macroeconomic variables.

B Tables

	L_{t-1}	S_{t-1}	C_{t-1}	CU_{t-1}	FFR_{t-1}	$INFL_{t-1}$	BSP_{t-1}	μ
L_t	0.990	0.019	0.000	-0.008	-0.008	0.034	-0.005	6.109
S_t	0.232	1.165	0.027	0.018	-0.229	-0.005	0.010	-2.384
C_t	-0.079	-0.110	0.881	0.013	0.124	-0.008	0.008	-1.475
CU_t	0.366	0.229	0.027	0.968	-0.246	-0.078	0.033	77.816
FFR_t	0.356	0.307	0.032	-0.001	0.664	-0.002	0.010	3.873
$INFL_t$	-0.014	-0.026	0.006	0.006	0.021	0.978	0.008	2.787
BSP_t	0.004	0.007	-0.014	-0.017	-0.004	-0.033	0.993	13.664

(a) Transition matrix Φ

	L_t	S_t	C_t	CU_t	FFR_t	$INFL_t$	BSP_t
L_t	0.068	-0.059	-0.005	0.007	0.004	-0.000	0.000
S_t		0.103	-0.020	0.021	0.020	0.005	-0.009
C_t			0.606	0.007	-0.007	0.003	0.022
CU_t				0.643	0.013	0.034	-0.027
FFR_t					0.030	0.004	-0.006
$INFL_t$						0.023	-0.003
BSP_t							0.095

(b) Variance-covariance matrix Q

Table 15: US YM4-KF, $\lambda = 0.035$

	L_{t-1}	S_{t-1}	C_{t-1}	GDP_{t-1}	MRO_{t-1}	$INFL_{t-1}$	EL_{t-1}	μ
L_t	0.921	-0.049	0.009	0.002	0.056	0.055	-0.044	3.239
S_t	0.144	1.100	0.023	0.026	-0.171	-0.033	0.055	-1.776
C_t	0.073	0.075	0.960	0.005	-0.053	-0.064	0.085	-2.200
GDP_t	-0.378	-0.417	-0.030	0.924	0.471	-0.027	0.060	3.168
MRO_t	0.129	0.102	0.020	-0.002	0.863	0.006	0.092	1.339
$INFL_t$	0.149	0.078	0.036	0.025	-0.127	0.894	0.130	1.617
EL_t	0.015	0.013	-0.009	-0.002	-0.011	-0.014	1.013	0.095

(a) Transition matrix Φ

	L_t	S_t	C_t	GDP_t	MRO_t	$INFL_t$	EL_t
L_t	0.041	-0.040	-0.007	0.013	0.005	-0.001	-0.000
S_t		0.063	-0.002	0.012	0.008	-0.001	-0.000
C_t			0.244	-0.015	-0.002	0.008	0.001
GDP_t				0.619	0.142	0.010	-0.003
MRO_t					0.049	0.001	-0.003
$INFL_t$						0.040	-0.000
EL_t							0.005

(b) Variance-covariance matrix Q

Table 16: EA, YM4-KF, $\lambda = 0.039$