UCLouvain
FUCaM MONS

LOUVAIN
School of Management

Louvain School of Management

Empirical Assessment of the Forecasting Performance and Robustness of the Dynamic Nelson-Siegel

A Two-Step vs. One-Step comparison

Author: Santiago Espitia
Supervisor: Frédéric Vrins
Academic year 2022-2023
Business Engineering, Major Financial Engineering

Abstract

The Dynamic Nelson-Siegel is a popular model for forecasting interest rates. However, the DNS has many variations and estimation procedures, it is not always clear which one is the most appropriate. Following an empirical out-of-sample approach, we first analyse various criteria for fixing the shape parameter in the context of the two-step estimation method of the DNS. We then compare the two-step and one-step approaches on market and simulated data. Lastly, we perform a robustness assessment of the two-step and one-step independent-factor DNS forecasting performance with respect to key parameters and starting values.

Acknowledgements

I would like to thank my thesis supervisor, professor Frédéric Vrins for his guidance and insight on many topics. His expertise allowed me to better understand many subjects and include many ideas on this work.

I also would like to thank Euroclear's team, Olga Reznikova, Mohit Lalatendu Mohanty and Giang Nguyen. Without their regular insight, guidance and feedback since the beginning of this work almost a year ago, this thesis would not have been possible.

Finally, I would like to thank my family for the effort of sending me abroad to study and for the support they were for me during all my studies despite the distance.

Contents

Nomenclature

t	Time index
τ	Maturity
h	Forecast horizon
T	Number of time steps
M	Number of maturities

y_t	$(M \times 1)$ Yield vector at time t
B	$(M \times 3)$ Measurement matrix
ϵ_t	$(M \times 1)$ Measurement error vector at time t
H	$(M \times M)$ Measurement error covariance matrix
f_t	(3×1) State vector at time t
A	(3×3) Transition matrix
η_t	(3×1) Transition error vector at time t
Q	(3×3) Transition error covariance matrix
μ	(3×1) Mean vector
P_t	(3×3) Error covariance at time t
$y_t(\tau)$	Interest rate for maturity τ in exercise at time t
$P_t(\tau)$	Discount factor at time t for a payment happening τ periods in the future
$f_t(\tau)$	Instantaneous forward rate for maturity τ at time t
l_t	Level factor at time t
s_t	Slope factor at time t
c_t	Curvature factor at time t
λ_t	Shape parameter at time t

1 Introduction

Interest rate forecasting is of paramount importance for a variety of applications including risk management, portfolio management, monetary policy, cash-flow discounting and the assessment of investment decisions. Many models have been developed to this purpose, with mixed results. One of these models is the Dynamic Nelson-Siegel (DNS). Due to its popularity, many variations and extensions have been introduced throughout the years.

Recently, the bankruptcy of several important banks, beginning by the collapse of Sillicon Valley Bank in march 2023, has shed light on the importance of proper interest rate risk management. In the context of banking regulation, the Pillar 2 [1] from the Basel framework evaluates risks and activities to ensure banks have enough capital on top of the minimum capital requirements. One of these risks is the interest rate risk in the banking book (IRRBB). Basel Committee on Banking Supervision, 2019 defines IRRBB as the current or prospective risk to the bank's capital and earnings arising from adverse movements in interest rates that affects the bank's banking book positions. Indeed, changes in interest rates impact the present value of the bank's assets, liabilities and off-balance sheet items and therefore impacts its economic value. In addition, changes in interest rates also impact interest-rate sensitive income, hence impacting the bank's earnings. The economic value of equity (EVE) and its derived measures (such as ΔEVE) are well known methods to measure IRRBB. The EVE is the net present value of all cash-flows originating from banking book assets, liabilities and off-balance sheet items. Naturally, to calculate the EVE one needs to discount the expected cash-flows of the portolio at the risk-free rate. With this context in mind, the DNS can be employed (and is often employed) to forecast the risk-free rate to be used in the EVE calculation. One can imagine that the accuracy of the forecasts can have profound impacts on the financial safety and profitability of these institutions.

While many studies have been conducted regarding the forecasting performance and robustness of the DNS model, this work makes several contributions to the literature.

[1] Pillar 1 sets the minimum capital requirements banks must hold to cover their credit, market and operational risks. Pillar 2 sets additional capital requirements on top of Pillar 1 requirements. Pillar 3 encourages transparency and market discipline among banks.

First, we compare the forecasting performance of the model between the DNS model under several variations of the two-step estimation method and the one-step estimation method, which is of interest given the many possible procedures available in the literature. Second, we perform this benchmark analysis with an European banking context in mind. This is of particular interest for the European banking sector because the particularities of the local market data can be determinant when it comes to choosing which procedure to use and how to calibrate your model. Third, we explore the sensibility of the DNS model to variations of the input parameters, helping practitioners to better understand the reliability and practical behaviour of the model. The main objectives of this work are the following: First, to theoretically describe the DNS model and its properties; Second, to empirically assess and compare the forecasting performance of the DNS following different estimation methods and different criteria for risk management purposes; Third, to assess the robustness of the DNS's forecasting performance with respect to variations in the input parameters.

In Section 2, we perform a literature review. We begin by diving into nature of interest rates and the yield curve. We then supply the necessary notation to understand the foundations of the original Nelson-Siegel model. After that, we introduce the original Nelson-Siegel model and we describe its theoretical properties, its behaviour and its interpretation. Then, we introduce the Dynamic Nelson-Siegel model, its alternative specifications and the different estimation methods we focus on.

In Section 3, we describe the data we dispose of for the empirical implementation. We also describe the forecasting methodology for the backtesting exercise and the methodology for the robustness analysis.

In Section 4, we present the results from our empirical exercise. We supply the comparative analysis between the different estimation methods, in addition to the comparative analysis between the alternative specifications of the model. Finally, we present the results from the robustness analysis.

In Section 5, we analyse the impact of the forecasting inaccuracies of our models in the EVE of several simulated portfolios.

In Section 6, we summarize our results, we interpret them and we discuss them. Finally, we end with a conclusion and a description of possible future work.

2 Literature review

2.1 On the nature of interest rates

Mishkin and Serletis, 2015 define interest rates as the cost of borrowing or the price paid for the rental of funds . There are many interest rates in the economy, e.g. rates of loans given by commercial banks, mortgage rates, interest rates of government and corporate bonds, and many more. However, from a theoretical standpoint, all interest rates are composed of the following five elements, as described in Basel Committee on Banking Supervision, 2016:

1. Risk-free rate: represents the theoretical interest rate an investor would expect from a completely risk-free investment for a particular maturity.
2. Market duration spread: premium added to compensate for the uncertainty endured by long-term held instruments. Indeed, their cash-flows become increasingly uncertain because the longer the instruments are held, the more they can be impacted by adverse changes in interest rates.
3. Market liquidity spread: premium added to compensate for the risk of not having an active market on which to do transactions when they are needed.
4. General market credit risk spread: premium required by market participants to compensate for a given credit quality. Indeed, the lower the credit quality, the lower the chances of paying back and the higher the compensation.
5. Idiosyncratic credit risk spread: premium representing the credit risk associated to the credit quality of the individual borrower and the specific instrument.

Because we are interested in the discounting of future cash-flows, we focus on forecasting the risk-free rate. From now on, we will also refer to the risk-free rate as simply the interest rate. Note that the risk-free rate should not be confused with the overnight rate and the policy rate. Indeed, the overnight rate, also called reference rate, is the interest rate to which participants of the overnight money market (such as banks) borrow and lend overnight funds to each other. According to Mishkin and Serletis, 2015, the overnight rate is the shortest term available and forms the base of the term-structure of interest rates. On the other hand, the policy rate is the Central Bank's target for the overnight rate. It is the main tool from the Central Bank to exert its monetary policy. Then again, the risk-free rate is a theoretical rate of a risk-free investment. Nonetheless, even if the risk-free rate is theoretical, it can be approximated by the interest rates of some nearly risk-free securities.

The relationship between interest rates and their maturities, also called the term structure of interest rates or the yield curve, is generally thought of as a bi-dimensional representation of risk-free rates for a set of maturities at a certain point in time. In other words, the yield curve refers to the term structure of interest rates of zero-coupon bonds without default risk, as explained by Nymand-Andersen, P., 2018. In the real world, interest rates are observable in discrete points for only a handful of maturities: there is no market data for every possible maturity. In addition, the yield curve is permanently changing through time, it is convenient then to add a temporal dimension to our conceptual framework, we can think about it as a collection of yield curves as they exist across time.

Yield curve modeling can follow different approaches. Nymand-Andersen, P., 2018 considers two families of models: first, models that make specific assumptions about changes in state variables and asset pricing methods using either equilibrium or arbitrage arguments; second, statistical models aiming to fit a smooth yield curve to the market data. The second family is divided into two sub-families: the spline-based methods and the parsimonious models, to which the Nelson-Siegel class of models belongs. For Nymand-Andersen, P., 2018, there are many theories underlying the term structure of interest rates , e.g. the liquidity preference theory, the pure expectations hypothesis, the segmented market hypothesis, the preferred habitat theory, among others. Nonetheless, our statistical approach doesn't make heavy use of them, so we consider they are out of the scope of this work. On the other hand, we do focus on the Nelson-Siegel model, which as we will see is not detached from financial theory.

2.2 Preliminary notation

Here we introduce the necessary notation and theoretical concepts to be able to understand the foundation of the Nelson-Siegel model. Let $P_t(\tau)$ be the present value of a Zero-Coupon Bond (also called ZCB or discount bond) with time to maturity τ seen from time t. In other words, it is the present value of 1 monetary unit receivable τ periods in the future. $P_t(\tau)$ is also called discount factor, and the graphical representation is called a discount curve. Indeed, the product of any future cash-flow by the respective discount factor would yield the present value of that cash-flow. Let $y_t(\tau)$ be the ZCB's yield to maturity at time t.

$$P_t(\tau) = e^{-\tau y_t(\tau)} \tag{1}$$

Let $f_t(\tau)$ be the instantaneous forward rate at time t.

$$f_t(\tau) = \frac{-P_t'(\tau)}{P_t(\tau)} \tag{2}$$

From these equations we can obtain the yield to maturity, which we also simply call yield.

$$y_t(\tau) = \frac{1}{\tau}\int_0^\tau f_t(u)du \tag{3}$$

2.3 The Nelson-Siegel model

Since its appearance in 1982, the Nelson-Siegel model and its variations have become increasingly widespread, to the point where many central banks and financial institutions use it to forecast the future shape of the yield curve. Even though the Nelson-Siegel model is able to forecast the yield curve of any instrument, we focus on the risk-free rate yield curve due to the risk management application we are considering.

Let's first exclude the temporal dynamics of the yield curve: we consider a cross-section or "slice" of the tridimensional yield curve, which corresponds to the traditional bi-dimensional yield curve at a certain point in time. Nelson and Siegel, 1987 assume that the forward rate $f(\tau)$ is the equal roots solution to a second order differential equation.

The expression of that solution being:

$$f(\tau) = \beta_1 + \beta_2 e^{\frac{-\tau}{\psi}} + \beta_3(\frac{\tau}{\psi} e^{\frac{-\tau}{\psi}}) \tag{4}$$

Where $\{\beta_1, \beta_2, \beta_3, \psi\}$ are the parameters of the model. Using the relationship between yield and forward rates in equation (3), we express this equation in terms of yields:

$$y(\tau) = \beta_1 + (\beta_2 + \beta_3)(\frac{1 - e^{\frac{-\tau}{\psi}}}{\frac{\tau}{\psi}}) - \beta_3 e^{\frac{-\tau}{\psi}} \tag{5}$$

Equation (5) is the original Nelson-Siegel model as proposed by its authors. However, from now on we are going to use an equivalent but slightly different formulation introduced in Diebold and Li, 2006. In this formulation, we regroup all coefficients, we define a new parameter $\lambda = \frac{1}{\psi}$ and we rename the coefficients $\beta_1, \beta_2, \beta_3$ as l, s, c. The Nelson-Siegel model becomes:

$$y(\tau) = l + s(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}) + c(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}) \tag{6}$$

The new parameters of the model are $\{l, s, c, \lambda\}$. We prefer the Diebold & Li formulation of the Nelson-Siegel model because the regressors (which we also call loadings) have less resembling shapes, allowing us to give them a certain interpretation. The parameters $\{l, s, c\}$ are also called factors. In Diebold and Li, 2006, the authors interpret the factors as the level, slope and curvature of the yield curve. Indeed, the loading of the level factor l remains constant in the short and long term, so we can interpret l as a long-term factor which mimics the level of the yield curve. The loading of the slope factor monotonically decreases to zero in the long-term, so we can think about s as a short-term factor resembling the slope of the yield curve. Finally, the loading of the curvature factor is low at short term maturities, it then increases in the medium term, and then tends to zero in the long-term, so we can think about c as a medium-term factor showing resemblance to the yield curve curvature. Many papers have visually demonstrated the resemblance of these factors and empirical proxies of the level, slope and curvature (e.g. Diebold and Li, 2006 and Diebold, F.X., Rudebusch, G.D. and Aruoba, B., 2006).

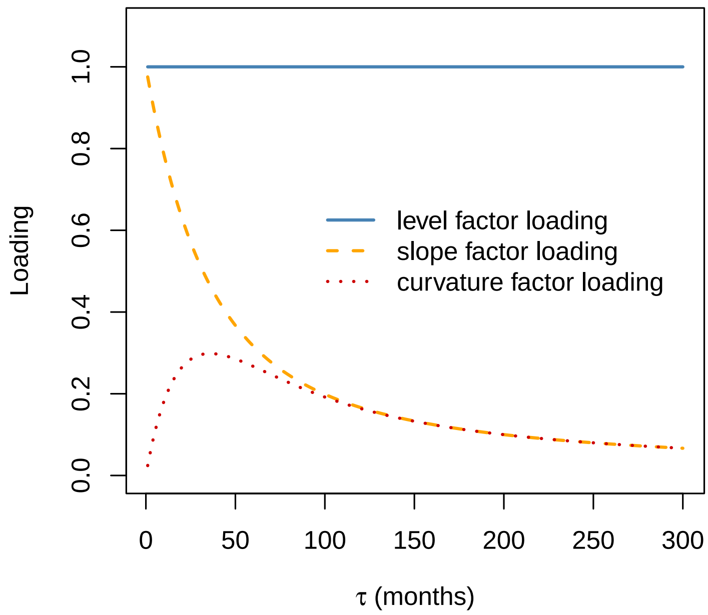


Figure 1: Loadings of the level, slope and curvature factors for $\lambda = 0.05$. Recall that the level factor loading is 1, the slope factor loading is $(1 - e^{-\lambda\tau})(\lambda\tau)^{-1}$ and the curvature factor loading is $(1 - e^{-\lambda\tau})(\lambda\tau)^{-1} - e^{-\lambda\tau}$.

We can extract even more information from the factors by looking at their sign. Indeed, the level factor l is straightforward: it is usually positive because the long term interest rate has almost always been positive. On the other hand, the slope factor s is often negative because the yield curve often has a positive slope. Lastly, the curvature factor c produces more curvature when it has the opposite sign of s. To have a better idea of how the factor signs influence the shape of the yield curve we can take a look at figure 2.

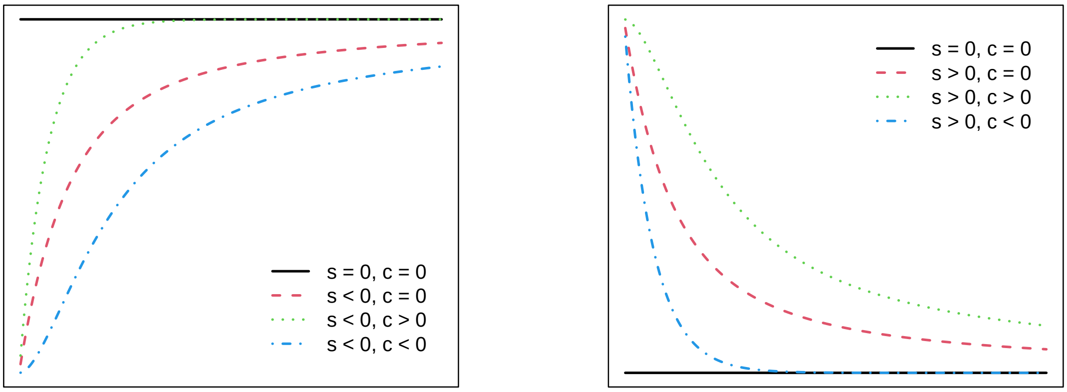
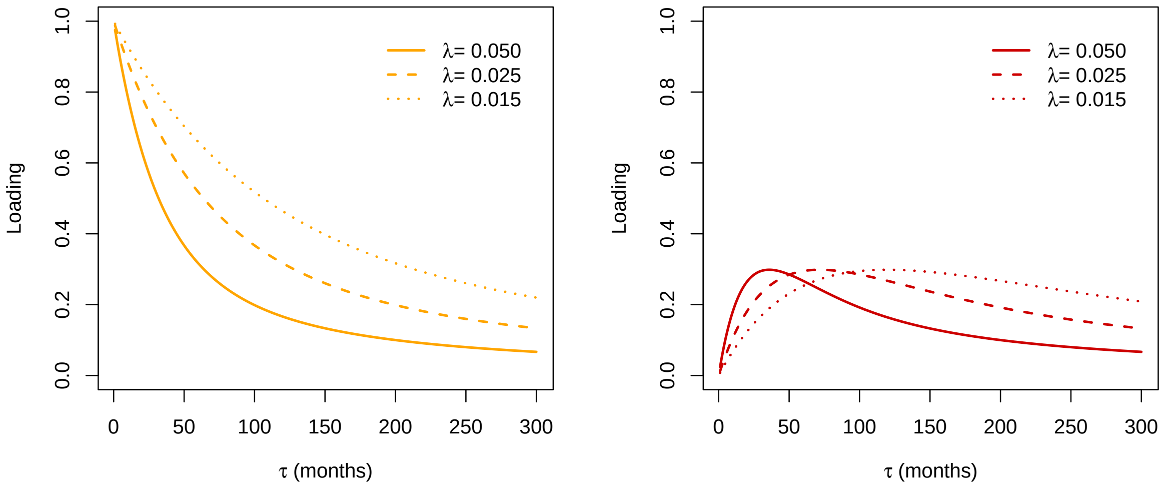


Figure 2: Illustrative yield curves produced with the Nelson-Siegel model. The level factor l and the shape parameter λ stay constant.

The parameter λ is often called exponential decay parameter or shape parameter. It governs the rate of decay of the slope and curvature loadings. High values of λ increase the speed of decay of the slope and curvature loadings, while small values decrease the rate of decay of those same loadings. In addition, the value of λ determines the location where the curvature loading will reach its maximum, meaning that it determines the location of maximum curvature on the fitted yield curve.



(a) Slope loadings

(b) Curvature loadings

Figure 3: Slope and curvature loadings for λ = 0.050, 0.025 and 0.015.

The Nelson-Siegel model exhibits several desirable properties. First of all, its parsimoniousness. Indeed, Diebold and Rudebusch, 2013 argue that the apparent high-dimensional nature of the yield curve seems to be well-approximated by a low-dimensional factor model. In addition, the relatively few explanatory variables in the Nelson-Siegel model make him better guarded against in-sample overfitting. It also produces smoother curves and promotes empirically tractable and trustworthy estimations. These are all useful advantages in our ultimate goal of forecasting the yield curve.

Secondly, the Nelson-Siegel model is flexible enough to reproduce a variety of yield curve shapes despite its parsimony. For instance, it can reproduce the traditional upward sloping, downward sloping, humped and inverted humped shapes. The precise shape is only dictated by the values of the parameters.

Thirdly, the Nelson-Siegel model enforces a number of theoretical principles of financial economics, as explained in Diebold and Rudebusch, 2013. Indeed, $lim_{\tau\to 0}\ P(\tau) = 1$ and $lim_{\tau\to\infty}\ P(\tau) = 0$, which are fundamental properties of the discount curve. The short and long ends of the term structure are respectively given by $lim_{\tau\to 0}\ y(\tau) = f(0) = l + s$ and $lim_{\tau\to\infty}\ y(\tau) = l$.

However, one may easily note that the Nelson-Siegel model is nonlinear relative to its parameters. Given this, it is possible to use nonlinear least squares to estimate the four parameters of the model. Another solution is to fix the parameter λ to impose the linearity of the model and then use ordinary least squares (OLS) regression to estimate the rest of the parameters.

2.4 The Dynamic Nelson-Siegel model

Let's now add back the temporal dimension of the yield curve, we represent the time axis by $t \in \{1, ..., T\}$. We consider henceforth a tri-dimensional representation of the yield curve. Diebold and Li, 2006 adapted the static Nelson-Siegel model, which is able to model a cross-section of the yield curve at a fixed point in time, into a dynamic version capable of modeling the yield curve through time. The following is the formulation of the

Dynamic Nelson-Siegel model.

$$y_t(\tau) = l_t + s_t(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}) + c_t(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}) \tag{7}$$

In the DNS model, the future shape of the yield curve is determined by the future values of the parameters $\{l_t, s_t, c_t, \lambda_t\}$. Meaning that to forecast the shape of the yield curve, one may forecast the dynamic parameters underlying it. We can also think about the dynamic parameters as variables with time series that we can forecast. For instance, the dynamics of the factors l, s and c are modeled as autoregressive processes in Diebold and Li, 2006. However, the nonlinearity of the DNS model produced by λ_t conditions the available methods that we can use to estimate it, as we will see later. It is therefore common procedure to turn λ_t into a constant. From now on, we note the constant shape parameter as λ.

It is worth noting that one theoretical drawback about the DNS is that it does not impose the necessary restrictions to prevent arbitrage, as noted in Filipović, 1999 and Christensen, J.H.E., Diebold, X. and Rudebusch. G.D, 2011. This limitation is tackled in other models derived from the DNS, such as the Arbitrage-Free Nelson-Siegel, discussed in Christensen, J.H.E., Diebold, X. and Rudebusch. G.D, 2011.

For practical reasons and conciseness, we translate the DNS model into a state-space framework. Let's add a stochastic error component ϵ to equation (7) and switch to matrix notation. The DNS model and the dynamics of the factors through time are respectively represented by two equations: the measurement equation and the transition equation. The measurement equation relates the current yields with the current factors while the transition equation relates the current factors with the past factors. Let $\{\tau_1, \tau_2, ..., \tau_M\}$ be the set of yield curve's maturities.

1) DNS model / measurement equation

$$y_t = Bf_t + \epsilon_t \tag{8}$$

where

$$y_t = \begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_M) \end{pmatrix}, \quad B = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_M}}{\lambda\tau_M} & \frac{1-e^{-\lambda\tau_M}}{\lambda\tau_M} - e^{-\lambda\tau_M} \end{pmatrix}, \quad f_t = \begin{pmatrix} l_t \\ s_t \\ c_t \end{pmatrix}, \quad \epsilon_t = \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_M) \end{pmatrix}$$

$$H = \begin{pmatrix} h_{11}^2 & 0 & \dots & 0 \\ 0 & h_{22}^2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & h_{MM}^2 \end{pmatrix}$$

We assume that the measurement error covariance matrix H is diagonal, implying that the errors on one maturity are independent from the errors on other maturities.

2) Factor's dynamic model / transition equation

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t \tag{9}$$

where

$$\mu = \begin{pmatrix} \mu^l \\ \mu^s \\ \mu^c \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \eta_t^l \\ \eta_t^s \\ \eta_t^c \end{pmatrix}$$

We model the factors as autoregressive processes. The specification of A and Q will determine whether we model the factor dynamics with an AR(1) or VAR(1) model, we describe later those two cases. We complete the representation of the model by specifying the covariance structure of the measurement and transition errors. We assume that the errors are normally distributed and that their covariance matrices are orthogonal to each other and to the initial state.

$$\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix}\right) \tag{10}$$

$$E(f_1\eta_t') = 0 \tag{11}$$

$$E(f_1\epsilon_t') = 0 \tag{12}$$

In equation (10), we state that both error vectors η_t and ϵ_t have zero mean and are not related with one another. Equations (11) and (12) state that the covariance between the error vectors and the initial factors is null [2]. In essence we consider that the errors don't depend on the value of the initial factors.

2.4.1 Independent-factor Dynamic Nelson-Siegel

In the case we specify A and Q as diagonal matrices we are implying that each factor evolves independently from the other ones: there is no cross-factor correlation. The Independent-factor Dynamic Nelson-Siegel (IDNS) is simply the DNS when the matrices A and Q are diagonal. Therefore, the IDNS may be represented as follows:

$$y_t = Bf_t + \epsilon_t \ , \quad \epsilon_t \ \sim \mathcal{N}(0, H) \tag{13}$$

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t \ , \quad \eta_t \ \sim \mathcal{N}(0, Q) \tag{14}$$

where

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

Note that equations (13) and (14) are the same as equations (8) and (9). Regarding the choice of the factor dynamics, AR(1) models are the simplest of all autoregressive models. This fact may prove useful in a forecasting context as it is well known that parsimonious models are less prone to overfitting. Despite its simplicity, the model can still provide accurate forecasts if its assumptions are satisfied.

[2] One can easily verify this by considering that $Cov(u, v) = E(uv')$ when $E(u)$ and/or $E(v)$ equal 0.

2.4.2 Correlated-factor Dynamic Nelson-Siegel

If we don't enforce the diagonality of A and Q, we are assuming that the value of each factor is influenced not only by past values of itself but by past values of the other factors, implying that there is cross-factor correlation. Consequently, the Correlated-factor Dynamic Nelson-Siegel (CDNS) is represented as follows:

$$y_t = Bf_t + \epsilon_t \ , \quad \epsilon_t \ \sim \mathcal{N}(0, H) \tag{15}$$

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t \ , \quad \eta_t \ \sim \mathcal{N}(0, Q) \tag{16}$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad Q = \begin{pmatrix} q_{11}^2 & q_{12} & q_{13} \\ q_{21} & q_{22}^2 & q_{23} \\ q_{31} & q_{32} & q_{33}^2 \end{pmatrix}$$

Note that equations (15) and (16) are the same equations as (8) and (9). Diebold and Li, 2006 suggest that a VAR model may perform poorer than a AR model in this context due to its potential for high in-sample overfitting as a consequence of its larger number of parameters. VAR models are more appropriate when there is high cross-factor correlation, which may not be our case. We still include it in our analysis because it is a variant that is not only often discussed and examined in research papers such as Diebold, F.X., Rudebusch, G.D. and Aruoba, B., 2006,Caldeira and al., 2016, Christensen, J.H.E., Diebold, X. and Rudebusch. G.D, 2011 and Huang, 2021, but has demonstrated to perform better than the IDNS in many of the previously mentioned studies.

2.4.3 Two-step estimation method

The simplest way to estimate the parameters of the DNS (including of course both the IDNS and CDNS) is the so-called two-step estimation method, or two-step approach. This approach was introduced in Diebold and Li, 2006, but it has been redesigned throughout the years. Currently, the two-step approach is more of a generalized procedure with many possible modifications than a well defined set of instructions. For our analysis, we selected several criteria present in the literature and we integrate them into our procedure in order to compare their adequacy regarding the improvement of the forecast performance of the DNS model. Note that the criteria are mutually exclusive: even if they are one after the other on the procedure below, we can only apply one. The next is the two-step approach we follow:

1. Select and fix the appropriate λ
   - Criterion 1: Select and fix the λ that minimizes the in-sample error between the empirical yield curve and a fitted yield curve. The error statistic can be the Mean Absolute Error (MAE) or the Root Mean Square Error (RMSE).
   - Criterion 2: Select and fix the λ that minimizes the correlation between the slope and curvature loadings.
   - Criterion 3: Select and fix the λ that maximizes the curvature of the estimated yield curve at the point of maximum curvature of the average empirical yield curve.
2. Perform a linear regression by OLS on the cross-section of the yield curve for each point in time to obtain the time serie of the factors. For each point in time, the factors are the coefficients we want to estimate, the loadings are the independent variables and the yields are the dependent variables.
3. Fit an autoregressive model to the estimated factors' time series.

In layman's terms, the two-step approach consists in estimating the static Nelson-Siegel for each point in time while maintaining a constant λ and then fitting an autoregressive model to the resultant time series of l, s and c. The two-step approach has the advantage of being simple, convenient and numerically stable. Indeed, it solely relies in OLS regression, which is quick, transparent and easy to perform. However, the two-step approach may be statistically suboptimal because the errors of the past time steps are ignored in the estimation of the next time step, as mentioned in Diebold and Rudebusch, 2013. The estimation for each point in time are made independently from the past estimation, as a consequence the procedure does not improve its inference capabilities based on past values. Below we discuss each of the criteria we consider.

Criterion 1: in-sample error minimization. The main idea behind this criterion is to select the λ that minimizes the in-sample differences between the empirical yield curve and a fitted yield curve. It is one of the approaches proposed by Ibáñez, 2015. Imagine you fix a certain value of λ, you can perform OLS regression at each time step to obtain a time serie for each factor. Using these estimated factors and the fixed value of λ, one can then fit a Nelson-Siegel yield curve to the empirical yield curve on each date just as you would for the two-step estimation method. However, instead of focusing on the factors, we focus on the fit of the yield curve with respect to the actual yield curve. An appropriate value of λ would allow the fitted curve to match as much as possible the shape of the actual yield curve. Note that we use the term *in-sample* error to avoid confusion with the *out-of-sample* error from the forecasts. Indeed, recall that we are not fitting any future yield curve, we are merely fitting the existing empirical yield curve and measuring the error of the fit. We use the MAE and the RMSE as our error statistics.

We can use numerical optimization methods to estimate the λ that minimizes the in-sample error. Based in our experience, it is nonetheless possible that there are local minima when the value of λ gets too further away from 0. For didactic purposes and even though it may not be our case, we use Brent's method to estimate the optimal shape parameter λ^* and avoid a hypothetical local optima by restricting λ to an upper and lower bound of 0.2 and 0 respectively. The optimization can be represented in the following way if we select the MAE as our error statistic:

$$\lambda^* = \underset{\lambda}{argmin} \frac{\sum_{t=1}^{T}\sum_{i=1}^{M} |y_t(\tau_i) - \hat{y}_t(\tau_i)|}{T \cdot M}$$

$$s.t. \quad 0 < \lambda < 0.2$$

On the other hand, it can be represented as follows if we choose the RMSE as our error statistic:

$$\lambda^* = \underset{\lambda}{argmin} \sqrt{\frac{\sum_{t=1}^{T}\sum_{i=1}^{M}(y_t(\tau_i) - \hat{y}_t(\tau_i))^2}{T \cdot M}}$$

$$s.t. \quad 0 < \lambda < 0.2$$

Where the time indexes $t \in \{1, ..., T\}$ represent the time steps of the yield curve. Note that $\hat{y}_t(\tau_i)$ is not a forecast since we are not forecasting the yield curve, but an in-sample fit at time t.

Criterion 2: slope and curvature loading correlation minimization. This criterion is also described in Ibáñez, 2015. The main idea is to select a λ that minimizes the correlation between the slope and curvature loadings. It is well known that multicollinearity between regressors difficults the accurate estimation of the coefficients in a linear regression. As shown by Annaert, J., Claes, A., De Ceuster, M., Zhang, H., 2013, high multicollinearity between the slope and curvature regressors produces unstable factor time series, which could negatively impacts our forecasts. While Annaert, J., Claes, A., De Ceuster, M., Zhang, H., 2013 use ridge regression to solve the multicolliearity problem, we use numerical optimization methods to find the value of λ that minimizes the square correlation between the slope and curvature loadings. Using the square correlation instead of the correlation is just an equivalent and more convenient problem setting. Indeed, by squaring the correlation we transform our problem to a mere minimization on which we can use basic optimization algorithms. We state the problem as:

$$\lambda^* = \underset{\lambda}{argmin}\ \rho^2_{L_s,L_c}$$

where

$$L_s = \begin{bmatrix} (1-e^{-\lambda\tau_1})(\lambda\tau_1)^{-1} \\ (1-e^{-\lambda\tau_2})(\lambda\tau_2)^{-1} \\ \vdots \\ (1-e^{-\lambda\tau_M})(\lambda\tau_M)^{-1} \end{bmatrix} \quad ; \quad L_c = \begin{bmatrix} (1-e^{-\lambda\tau_1})(\lambda\tau_1)^{-1} - e^{-\lambda\tau_1} \\ (1-e^{-\lambda\tau_2})(\lambda\tau_2)^{-1} - e^{-\lambda\tau_2} \\ \vdots \\ (1-e^{-\lambda\tau_M})(\lambda\tau_M)^{-1} - e^{-\lambda\tau_M} \end{bmatrix}$$

L_s and L_c are simply vectors of the slope and curvature loadings for each maturity and $\rho^2_{L_s,L_c}$ is the squared Pearson correlation between L_s and L_c. Ibáñez, 2015 uses vectors of daily hypothetical maturities for each loading, however we use vectors only based on the maturities available to us. We do this because the correlation between the two loadings depends on the maturity vector (one could quickly verify this by plugging different maturity vectors for the loadings and calculating their correlation). We could even end up increasing the correlation between our two loadings if we select a λ based on the correlation of a different maturity vector. Therefore, we ensure the least correlation between our loadings by using only the maturities we have and not any hypothetical maturity. The slope and curvature loading correlation minimization criterion is more universal than the other criteria in the sense that it does not depend on the data: it only depends on the maturity vector. One can input a different yield curve and use the same value of λ as long as it maintains the same maturity vector.

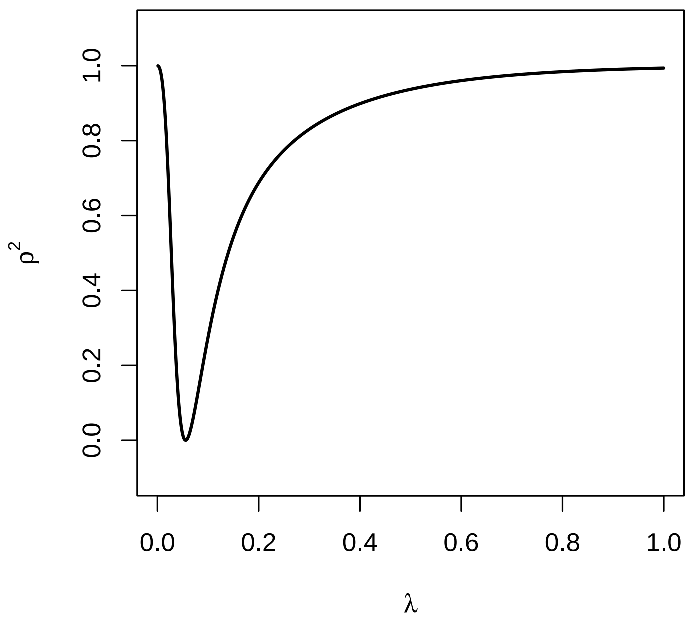


Figure 4: Squared correlation between the slope and curvature loadings as a function of λ.

As shown in figure 4, our objective function has one global minima and no local minima. The optimization can be easily performed.

Criterion 3: point of maximum average curvature. This is the criterion used by Diebold and Li, 2006. The main idea is to select a value of λ such that the curvature loading reaches its maximum at the same maturity the average empirical yield curve reaches its maximum curvature. In other words, we want to select a λ such that the curvature of the fitted yield curves matches as much as possible the curvature of the empirical yield curves. Naturally, different yield curves can have different locations for their maximum curvature, it may vary from country to country and from instrument to instrument. For some curves, it may be fairly easy to visually determine the point of maximum curvature (for instance, in the average yield curve from Diebold and Li, 2006 it is easy to see the point of maximum curvature, which is set by Diebold & Li at 30 months). However, the discrete nature of the yield curve can make it trickier to see the point of maximum curvature, not to mention that some yield curves can be near flat. A solution can be to fit a smooth curve and then calculate the point of maximum curvature, but even this leaves room for arbitrariness because the modeler still can decide how curvy its fitting curve should be. Once you know the maturity at which the curvature loading should reach its maximum, one can proceed by using numerical methods to estimate the value of λ^*.

2.4.4 One-step estimation method

The alternative estimation method we analyse is the one-step estimation method, also called one-step approach. Introduced by Diebold, F.X., Rudebusch, G.D. and Aruoba, B., 2006, the one-step approach is the other general procedure used to estimate the DNS model. As discussed by Diebold and Rudebusch, 2013, the one-step approach refers to a general procedure that may be extended and modified with different tools. The main idea behind it is that the state-space formulation of the DNS model allows for a simultaneous estimation of the parameters, usually though Maximum Likelihood Estimation (MLE), the Kalman filter and an optimization algorithm [3]. We focus on the traditional approach of the one-step estimation method, which is referred by Diebold and Rudebusch, 2013 as Numerical Gaussian Quasi Maximum Likelihood Estimation (Numerical Gaussian QMLE in short). Our one-step approach is the following:

1. Estimate all parameters at once following a MLE criteria via numerical optimization:
   - Give the initial guesses for H, Q, λ, A and μ.
   - Give the initial values for the factors' time series f_1 and for the initial error covariance matrix P_1.
   - Define the objective function $-l(y_1, ...y_T; \Theta)$ as the negative of the logarithm of the Gaussian pseudo-likelihood produced by a run of the Kalman filter.

[3] The Berndt–Hall–Hall–Hausman algorithm is used in Diebold, F.X., Rudebusch, G.D. and Aruoba, B., 2006, the Nelder-Mean algorithm is used in Christensen, J.H.E., Diebold, X. and Rudebusch. G.D, 2011, while the Broyden–Fletcher–Goldfarb–Shanno algorithm is used in other implementations.

- Set non-negativity constraints for λ and each element in H and Q. In addition, restrict elements in A to be inside the unit circle to ensure the stationarity of the autoregressive process.
- Optimize the objective function by varying the parameters each iteration until convergence.

2. Once the optimization has converged, run once the Kalman filter with smoothing with the optimal parameters to obtain the estimated factors' time series. Then extract the last value of each factor.

3. Fit an autoregressive model using the optimal parameters beginning from the factor's last values.

In layman's terms, the Kalman filter is run with the current parameter values for each iteration of the optimization. The outcome of each iteration is a negative log-likelihood estimate, which the optimizer tries to minimize by varying the parameters. We use the Nelder-Mead algorithm [4] to minimize the negative log-likelihood at each iteration. The constraints are implemented by the means of logarithmic barrier functions. Nonetheless, the constraints change slightly depending on whether we want to estimate the IDNS or CDNS model. We implement the one-step approach by the means of the R packages [5] **kalmanfilter** and **maxLik**. The one-step approach for estimating the IDNS model can be represented as follows:

$$\Theta^*_{IDNS} = \underset{\Theta_{IDNS}}{argmin} \ -l(y_1, ..., y_T; \Theta_{IDNS}) \tag{17}$$

$$s.t. \quad \lambda > 0$$

$$-1 < a_{ii} < 1 \qquad \forall i \in \{1, 2, 3\}$$

$$h^2_{ii} > 0 \qquad \forall i \in \{1, ..., M\}$$

$$q^2_{ii} > 0 \qquad \forall i \in \{1, 2, 3\}$$

Θ^*_{IDNS} is the set of optimal parameters of the IDNS. In our case, there are 31 pa-

[4]We also tried the BFGS algorithm but the Nelder-Mead yielded better results.

[5]The R package **Fast Kalman Filter** and an in-house implementation did not produce satisfactory results.

rameters to be optimized in the IDNS model (λ, 3 from the diagonal of A, 15 from the diagonal of H, 3 from the diagonal of Q, 3 from μ, 3 from f_1 and 3 from the diagonal of P_1). On the other hand, the one-step approach for estimating the CDNS model can be represented in the following way:

$$\Theta^*_{CDNS} = \underset{\Theta_{CDNS}}{argmin} \; -l(y_1, ..., y_T; \Theta_{CDNS}) \tag{18}$$

$$s.t. \quad \lambda > 0$$

$$-1 < a_{ij} < 1 \qquad \forall i, j \in \{1, 2, 3\}$$

$$h^2_{ii} > 0 \qquad \forall i \in \{1, ..., M\}$$

$$q^2_{ii} > 0 \qquad \forall i, j \in \{1, 2, 3\}$$

Θ^*_{CDNS} is the set of optimal parameters of the CDNS. On its foundations, the CDNS optimization has essentially the same specifications as the IDNS optimization, with the difference that for the CDNS we want every non-diagonal and diagonal element in the matrix A to be comprised inside the unit circle, instead of just the diagonal ones. In our case, our CDNS model has 40 free parameters (λ, 9 from A, 15 from the diagonal of H, 6 from Q, 3 from μ, 3 from f_1 and 3 from the diagonal of P_1). Note that we don't need to estimate all the 9 elements of the matrix Q due to the symmetric properties of covariance matrices: we simply need to estimate the lower (or upper) triangle and then project it into the other triangle.

Regarding the choice of the initial guesses we need to supply in steps 1 and 2 of the one-step approach, most of them can be extracted from the results obtained via the two-step approach. Indeed, λ was fixed on the first step of the two-step approach. The initial guess for the matrix H can be estimated by calculating the variances of the errors between the empirical yield curve and the fitted yield curve of the two-step DNS. The guess for the matrix A can come from the estimated autoregressive coefficients of the two-step DNS. The guess for matrix Q can come from the estimated error covariance matrix of the autoregressive model fitted with the two-step DNS. The guess for the vector μ can be the unconditional mean of the factors deduced from the estimated intercept of the autoregressive model fitted with the two-step DNS. The guess for the initial factors f_1

can be the factors found in the first time step of the two-step DNS. Finally, the matrix P_1 can be the unconditional variance of the factors using the parameter matrices found by the two-step DNS, note however that we later only take the diagonal of P_1 as parameters. The unconditional mean and unconditional variance of the factors, μ and Σ, are given by:

$$\mu = (I_3 - A)^{-1} c \tag{19}$$

$$vec(\Sigma) = (I_9 - A \otimes A)^{-1}\, vec(Q) \tag{20}$$

Where c is the vector of intercepts of our autoregressive model, I_k is a (k $\times$ k) identity matrix, $vec(\cdot)$ denotes the vectorization operation and $\otimes$ referes to the Kronecker product. The main advantage of the one-step estimation is that it estimates all the parameters of the model at once, which is better from an inferential point of view. Recall that the two-step approach lacked in this regard because each cross-sectional yield curve has to be estimated independently from the other ones at each point in time in a sort of step by step fashion. Despite this theoretical advantage, the one-step approach relying on traditional gradient-based algorithms has practical issues regarding the accuracy of the results. As mentioned in Diebold and Rudebusch, 2013, the gradient-based one-step approach not always converges in a trustworthy way. Even if our Nelder-Mead algorithm is not gradient-based, our results could also lack trustworthiness due to the fundamental resemblance of the one-step procedure in both cases. In addition, the procedure is slower and more complex than the two-step approach.

We give now a brief description of how the Kalman filter works. As explained in Kim and Bang, 2018, Kalman filters are algorithms used to estimate states based on linear dynamical systems in state space format. It uses a recursive procedure. Let the superscript $-$ denote the prior (or predicted) and $+$ denote the posterior (or updated) estimates. The hat operator $\hat{}$ denotes an estimate of the actual variable, for instance $\hat{f}_t^+$ is the posterior estimate of f_t. In our case, the Kalman filter algorithm takes as input the parameters H, Q, λ, A, μ, the initial values for the factor's time series $\hat{f}_1^+$, the initial error covariance matrix P_1^+, and finally the observed yields y_t. It produces an estimate of the factor's time series $\hat{f}_t^+$. The algorithm has two stages: prediction and update.

Prediction stage

- Predicted state estimate

$$\hat{f}_t^- - \mu = A(\hat{f}_{t-1}^+ - \mu) \tag{21}$$

- Predicted error covariance

$$P_t^- = AP_{t-1}^+ A' + Q \tag{22}$$

Update stage

- Measurement residual

$$e_t = y_t - B\hat{f}_t^- \tag{23}$$

- Kalman gain

$$K_t = P_t^- B'(H + BP_t^- B')^{-1} \tag{24}$$

- Updated state estimate

$$\hat{f}_t^+ = \hat{f}_t^- + K_t e_t \tag{25}$$

- Updated error covariance

$$P_t^+ = (I - K_t B)P_t^- \tag{26}$$

$\hat{f}_1^+$ and P_1^+ are the initialization values. The Gaussian log-likelihood function can be obtained via prediction error decomposition. The log-likelihood function is:

$$\log l(y_1, ..., y_T; \Theta) = \sum_{t=1}^{T} (-\frac{M}{2}\log(2\pi) - \frac{1}{2}\log|\Omega_t| - \frac{1}{2}e_t' \Omega_t^{-1} e_t)$$

Θ is the set of model's parameters, $\Omega_t = H + BP_t^- B'$ and $|\cdot|$ denotes the determinant. As explained in Kim and Bang, 2018, the Kalman filter provides optimal estimates as long as the necessary assumptions are met. The main assumptions underlying the algorithm are that measurement models are linear, meaning that they can be expressed with the matrices A and B. In addition, the transition and measurement noise are additive Gaussian.

For each iteration, the prediction stage of the Kalman filter aims to predict both the state variables (our factors) using our autoregressive model in equation (21) and the error covariance in equation (22). The predicted state estimate is simply the expected value at time t given the factors at time $t-1$ and the specified autoregressive model. The predicted error covariance at time t, which we can think about as the uncertainty at time t given the accumulated uncertainties of all previous steps, is the uncertainty of the value of the state variable at time $t-1$ in addition to the uncertainty of step t. Regarding the update step, the objective is to combine the information from the measurements and from the predictions. In order to combine both information, the updated state estimate and the updated error covariance are weighted averages of the information coming from the measurement and from the predictions. The Kalman gain calculated in equation (24) is the weight that determines how much more a source is trusted over the other. The Kalman gain is the proportion of the error covariance with respect to the sum of the error covariance plus the measurement error, which is essentially how much the uncertainty of the state variables represents with respect to the total uncertainty at that step. If the current state uncertainty represents a big part of the total uncertainty, then the Kalman gain will tend to 1 and will give more weight to the measurement information in equation (25).

3 Data and methodology

3.1 The data

Our market data consists of an end-of-month risk-free yield curve for the period going from February 28 2007 to December 03 2022, giving a total of 191 observations across the time axis. The yield curve was bootstrapped from the Euro Overnight Index Swap (OIS) curve obtained from Bloomberg (see appendix 8.2 for more details). For most of the time, the floating leg of the EUR OIS paid the EONIA rate. However, since 2019 the ESTRON rate has progressively replaced the EONIA. Our yield curve has 15 maturities: 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144 and 180 months.

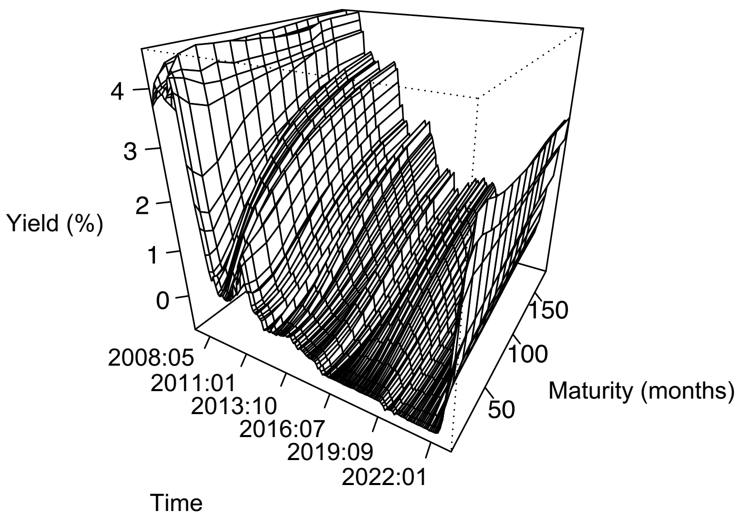


Figure 5: Bootstrapped risk-free yield curve from February 2007 to December 2022. There are 191 observations across the time axis.

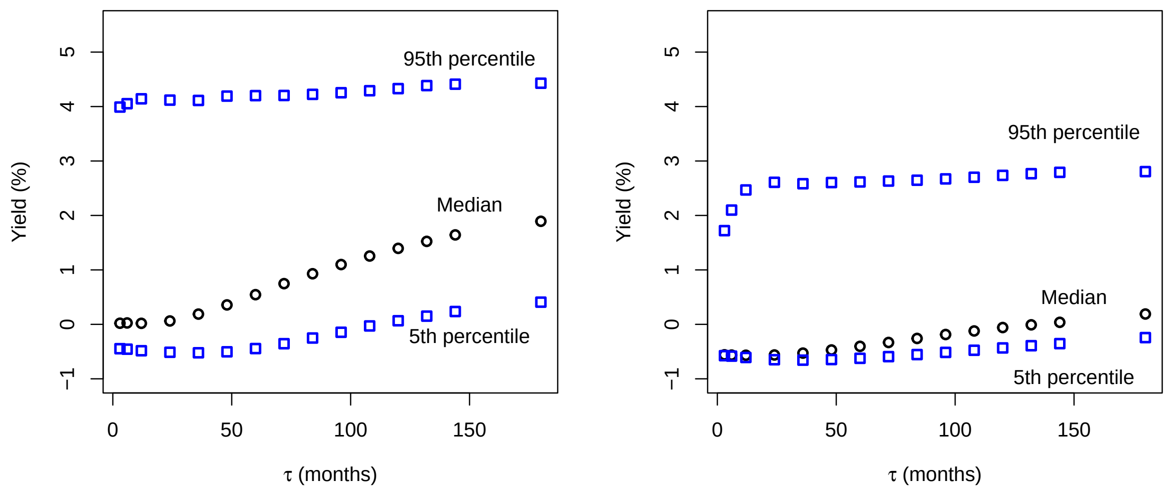
For our empirical exercise, we divide our data into a training set and a test set. Following an allocation of 80/20, the training set contains 152 dates or 152 cross-sections of the yield curve, while the test set contains 39. The training set yield curve covers the period from February 2007 to September 2019 while the test set yield curve covers from October 2019 to December 2022. Below are the descriptive statistics of the training and test set:

Maturities	Mean	Std.Dev	Min	Max	Skewness	Kurtosis	ACF(1)	ACF(6)	ACF(12)	PACF(1)	PACF(6)	PACF(12)
3	0.57	1.43	-0.59	4.25	1.71	4.49	0.91	0.78	0.53	-0.10	-0.03	-0.02
6	0.58	1.45	-0.64	4.32	1.71	4.49	0.91	0.77	0.53	-0.09	-0.01	0.03
12	0.61	1.46	-0.71	4.54	1.66	4.42	0.90	0.77	0.54	-0.09	0.02	0.08
24	0.70	1.46	-0.78	4.62	1.46	3.94	0.90	0.78	0.59	-0.07	0.04	0.08
36	0.83	1.47	-0.81	4.59	1.25	3.39	0.91	0.80	0.64	-0.06	0.06	0.07
48	0.98	1.47	-0.81	4.57	1.07	2.98	0.92	0.81	0.66	-0.06	0.07	0.07
60	1.12	1.47	-0.79	4.52	0.92	2.66	0.92	0.82	0.68	-0.05	0.07	0.06
72	1.27	1.46	-0.76	4.51	0.80	2.43	0.92	0.83	0.69	-0.03	0.07	0.06
84	1.40	1.44	-0.72	4.51	0.70	2.26	0.92	0.83	0.70	-0.02	0.06	0.05
96	1.53	1.43	-0.67	4.52	0.62	2.15	0.92	0.84	0.70	-0.01	0.06	0.04
108	1.64	1.41	-0.61	4.54	0.57	2.07	0.92	0.84	0.71	-0.00	0.05	0.04
120	1.75	1.40	-0.56	4.56	0.52	2.01	0.92	0.84	0.71	0.01	0.05	0.03
132	1.84	1.39	-0.52	4.58	0.48	1.97	0.92	0.84	0.71	0.01	0.04	0.03
144	1.93	1.38	-0.47	4.60	0.44	1.93	0.92	0.84	0.71	0.02	0.04	0.03
180	2.11	1.34	-0.34	4.61	0.36	1.86	0.92	0.83	0.71	0.03	0.04	0.02

Table 1: Training set yield curve descriptive statistics. Yield curve sample from February 2007 to September 2019. This curve has 152 observations.

Maturities	Mean	Std.Dev	Min	Max	Skewness	Kurtosis	ACF(1)	ACF(6)	ACF(12)	PACF(1)	PACF(6)	PACF(12)
3	-0.26	0.75	-0.59	2.25	2.40	7.33	0.43	0.02	-0.06	-0.15	0.03	-0.01
6	-0.18	0.89	-0.61	2.67	2.22	6.46	0.49	0.09	-0.06	-0.14	0.06	-0.03
12	-0.07	1.04	-0.64	3.12	2.00	5.57	0.56	0.18	-0.06	-0.12	0.13	-0.06
24	0.02	1.10	-0.67	3.14	1.70	4.49	0.62	0.31	-0.06	-0.10	0.16	-0.09
36	0.07	1.09	-0.66	3.02	1.55	4.02	0.65	0.36	-0.05	-0.10	0.17	-0.11
48	0.10	1.09	-0.65	2.95	1.48	3.77	0.67	0.38	-0.05	-0.10	0.17	-0.12
60	0.14	1.08	-0.63	2.92	1.44	3.64	0.68	0.39	-0.04	-0.10	0.17	-0.12
72	0.18	1.07	-0.60	2.90	1.41	3.55	0.68	0.40	-0.04	-0.10	0.17	-0.12
84	0.23	1.06	-0.57	2.87	1.38	3.46	0.69	0.40	-0.04	-0.10	0.17	-0.11
96	0.28	1.05	-0.53	2.87	1.36	3.39	0.69	0.40	-0.03	-0.10	0.17	-0.11
108	0.33	1.05	-0.49	2.88	1.33	3.33	0.70	0.41	-0.03	-0.10	0.17	-0.10
120	0.38	1.04	-0.44	2.90	1.32	3.28	0.70	0.41	-0.03	-0.10	0.17	-0.10
132	0.43	1.04	-0.40	2.91	1.30	3.23	0.70	0.41	-0.03	-0.11	0.17	-0.09
144	0.47	1.03	-0.36	2.92	1.29	3.19	0.70	0.41	-0.03	-0.12	0.16	-0.09
180	0.57	0.99	-0.26	2.88	1.25	3.09	0.71	0.42	-0.02	-0.13	0.16	-0.08

Table 2: Test set yield curve descriptive statistics. Yield curve sample from October 2019 to December 2022. This curve has 39 observations.



(a) Training set (b) Test set

Figure 6: Median, 5th and 95th percentiles of the training set and test set yield curve.

3.2 Forecasting backtest methodology

We perform an out-of-sample rolling origin forecast with constant holdout and expanding training set (see appendix 8.1 to have an illustration of this type of forecast). At the initial step, we use all the training set to calibrate the model and produce our forecasts. Simulating the passage of time, on the next step a new cross-section of the yield curve is (so called) observed and incorporated into the training set. The model is then calibrated with this augmented training set and new forecasts are produced. This process is repeated until we reach the end of the test set. Our forecast horizon will be of 12 months, meaning that at each step we produce forecasts for 1, 2, ..., up to 12 months ahead. Note that we produce forecasts for each of the maturities of our training set yield curve. At each step, we produce 180 individual forecasts (15 maturities $\times$ 12 months ahead). Having 28 steps, we generate a total of 5040 individual forecasts per model on a single backtest run. On the last step, the model is calibrated with a training set of 179 dates (or cross-sectional yield curves) due to each new yield curve being absorbed by the training set. The interest rate forecasts of the IDNS and CDNS are given by:

$$\hat{y}_{t+h|t}(\tau) = \hat{l}_{t+h|t} + \hat{s}_{t+h|t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \hat{c}_{t+h|t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{27}$$

Or equivalently, expressed in a state-space framework.

$$\hat{y}_{t+h|t} = B\hat{f}_{t+h|t} \tag{28}$$

Where the forecasted underlying factors are given by.

$$(\hat{f}_{t+h|t} - \mu) = A(\hat{f}_{t+h-1|t} - \mu) \tag{29}$$

To quantify the performance (or lack of performance) of the competing models, we select the MAE and the RMSE as our error statistics. We calculate the out-of-sample MAE and RMSE of the predicted yields for each maturity and for a set of $h \in \{1, 3, 6, 12\}$ months (or periods) ahead. Let $\hat{y}_{t+h|t}(\tau)$ be the forecasted interest rate for maturity τ to be exercised at time $t + h$. The MAE and RMSE across the time indexes $t \in \{1, ..., T\}$ for maturity τ and forecasting horizon h are given by:

$$MAE_{\tau,h} = \frac{\sum_{t=1}^{T} |y_{t+h}(\tau) - \hat{y}_{t+h|t}(\tau)|}{T}$$

$$RMSE_{\tau,h} = \sqrt{\frac{\sum_{t=1}^{T} (y_{t+h}(\tau) - \hat{y}_{t+h|t}(\tau))^2}{T}}$$

Note that the time indexes go from the first date of the test set yield curve up to the last date. In addition, $y_{t+h}(\tau)$ is not a Nelson-Siegel produced yield, but an actual yield to be in exercise at time $t + h$.

We perform two forecasting competitions. First, we compare the out-of-sample MAE and RMSE of the forecasts produced by the two-step IDNS following the 3 different criteria for selecting the value of the shape parameter λ. We perform the same comparison for the two-step CDNS. The objective is to determine which of the three criteria for selecting the parameter λ maximizes the forecasting performance of the two-step DNS.

Second, we compare the most performing two-step DNS models (both the previous winners of the IDNS competition and the CDNS competition) with the one-step IDNS and CDNS. We also include the well known Random Walk model as a benchmark for all models in this second competition. The Random Walk model forecast is simply:

$$\hat{y}_{t+h|t}(\tau) = y_t(\tau) \tag{30}$$

The second competition has two objectives. First of all, to determine which one between the two-step DNS and the one-step DNS has better forecasting performance. Second of all, to determine if the DNS model is a competitive alternative to the benchmark Random Walk model, which postulates that the best forecast of the yield is simply the last known value of the yield.

In our EVE forecasting exercise, we extend our backtesting approach by simulating various portfolios and forecast their EVE based on the forecasted interest rates of each model. Then, we compare those forecasted EVE values with the actual EVE corresponding to those periods. At each step of the backtest, we consider ourselves in the present (h = 0) and issue forecasts for $h \in \{1, 3, 6, 12\}$ periods ahead. We calculate the actual EVE at time t as follows:

$$EVE_t = \sum_{i=1}^{M} P_t(\tau_i) \times CF_{t,\tau_i} \tag{31}$$

Where $P_t(\tau_i)$ is the discount factor of maturity index τ_i at time t and CF_{t,τ_i} is the portfolio cash-flow to be cashed in (or out) τ_i periods after time t. The forecasted EVE to be in effect at time $t + h$ is given by:

$$\widehat{EVE}_{t+h|t} = \sum_{i=1}^{M} \hat{P}_{t+h|t}(\tau_i) \times CF_{t+h|t,\tau_i} \tag{32}$$

Note that in this exercise, we first assume that the cash-flows of our portfolio stay constant through time; second, we assume that the cash-flows are known in advance with absolute certainty, hence we don't represent them as forecasts. As a consequence of these assumptions, the discrepancies between the actual and forecasted EVE can only come from discrepancies between the actual and forecasted interest rates.

3.3 Robustness analysis methodology

We evaluate the robustness of the DNS model in two aspects. First, we perform a sensitivity analysis of the two-step IDNS out-of-sample forecasting performance with respect to the shape parameter λ. Given the many ways the parameter λ can be selected and fixed, our aim is to assess how its value impacts the forecasting performance of the model. Due to the higher popularity of the IDNS compared to the CDNS among practitioners, we only examine the IDNS. In order to carry the analysis, we generate a grid of values of λ. For each value, we perform a backtest with the IDNS and the particular λ as input in all of the backtest steps. This is different from the previous approach in which we re-calibrate the value of λ on each step of the backtest to simulate a real-life application. In other words, here there is no re-calbration: the value of λ stays constant. In virtue of the quick running time of the two-step IDNS, we are able to run many backtests in a reasonable amount of time and obtain results for a full range of values of λ.

Second, we perform a sensitivity analysis of the out-of-sample forecasting performance of the one-step IDNS with respect to the initial guesses of the parameters the user has to supply. It has been noted by Diebold and Rudebusch, 2013 that the one-step DNS may produce unstable and unreliable estimates. We aim to determine if this issue is caused by a high degree of sensibility to the starting points of the optimization. If the optimization is too sensible to the initial guesses (the starting values of the optimization) then more care should be put in selecting appropriate starting values or, in the most extreme case, avoid using this type of algorithms and lean towards more robust alternatives (Diebold and Rudebusch, 2013 discuss alternative algorithms for the one-step approach). Due to the high computational burden of the one-step approach, we are not able to run backtests for a whole set of values for each of the 31 parameters [6]. Instead, we apply a series of downward and upward shocks to the matrices and vectors containing the parameters. For instance, we apply a downward and upward shock of x% to the initial parameters $A, Q, H, \mu, \lambda, f_1$ and P_1 to be supplied as starting values for the MLE optimization at each iteration of the backtest. Note that we apply the shocks in a separate way: we don't apply a shock to two parameters at the same time. Then we extract and compare the MAE of the model

[6]The runtime of a single iteration inside a backtest is about 3 minutes with our 2.4 GHz Intel Core i5 processor. 10 values per parameter would take us about 18 days of runtime.

under shocks with respect to the model without shocks. A small difference in MAE would indicate robustness with respect to the guesses of the initial parameters. Contrariwise, a big difference in MAE would indicate that the one-step estimation method is sensible to the choice of parameter guesses. With this approach, we can carry through our analysis within a reasonable amount of time while maintaining enough granularity to determine which matrices and vectors of initial parameter guesses the model is most sensible to.

4 Results

4.1 Two-step estimation method: empirical assessment

We follow the three criteria and calculate the values of λ^*.

Criterion 1: in-sample error minimization. Regrading the in-sample error minimization criterion, one can appreciate in figure 7 how the objective functions have only one global minimum and no local minima. This naturally simplifies our task to find the value of λ at the global minimum. Following this criterion, the calculated value of λ^* given the initial training set is 0.0288[7] if we minimize the MAE and 0.0308 if we minimize the RMSE. By *initial training set* we mean that the values of λ^* are calculated based on the yield curve of the first step of the backtest. Recall that in a backtest run, this procedure has to be repeated each iteration with an expanded training set. Consequently, at each iteration we would obtain a slightly different value of λ^* (see figure 10 to appreciate how the values of λ^* change at each iteration).

[7] We work with monthly maturities. If we were to convert them to yearly maturities we would have calculated $\lambda^* = 0.3458$ for the in-sample MAE minimization and 0.3693 for the in-sample RMSE minimization.

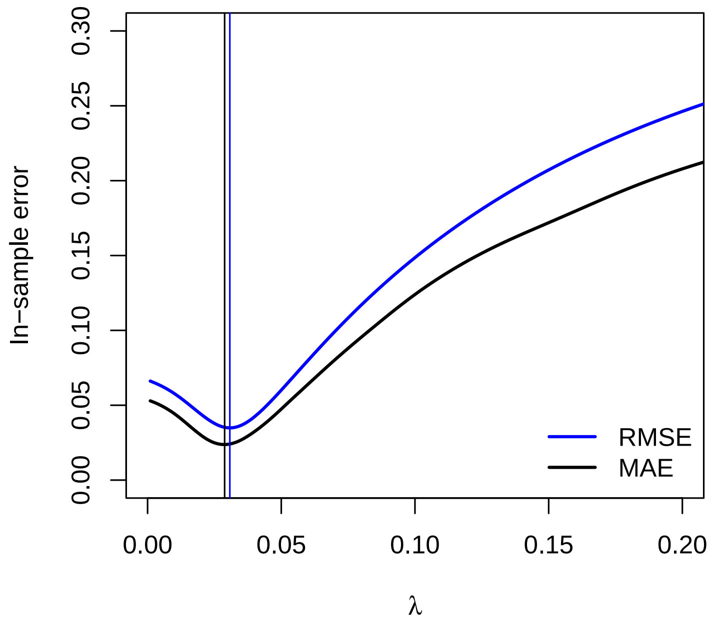


Figure 7: In-sample MAE and RMSE between the initial training set yield curve and a fitted yield curve produced with different values of λ. The vertical lines represent the values of λ^* relying on the MAE and RMSE minimization. *Note*: The curves and the values of λ^* depicted here correspond to the initial training set yield curve, meaning that they correspond to the first step of the backtest. In each subsequent step of the backtest a re-calibration is made with new data.

Criterion 2: slope and curvature loading correlation minimization. The value of λ^* obtained through this criterion does not change according to the data: it only changes according to the maturity vector. In our case, we calculated a value of 0.0557 [8] for λ^*. Contrary to criterion 1, this value stays constant throughout the backtest.

Criterion 3: point of maximum average curvature. One can observe in figure 8 that the point of maximum curvature in the training set yield curve is not as apparent as one would ideally hope. Nonetheless, we choose the so mentioned point to be at a maturity of 30 months. There are two reasons for that. First, there is an observable curvature at that point. Second, 30 months is also the point of maximum curvature chosen by Diebold and Li, 2006 and other applications. To calculate the value of λ^*, one can simply calculate the value that maximizes the curvature loading at 30 years using numerical methods. Since the objective function (which is the curvature loading) is a function with one maximum

[8] If we were to use yearly maturities instead of monthly, we would have calculated $\lambda^* = 0.6692$.

and no local maxima, it is a simple optimization to implement. The estimated value of λ^* following this criterion is 0.0597 [9]. In a backtest run, this value stays constant iteration after iteration since we assume the point of maximum curvature stays the same.

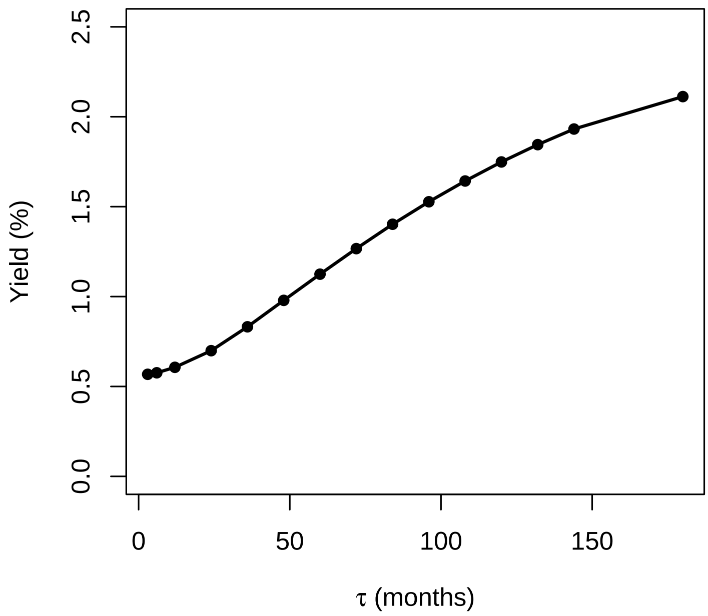
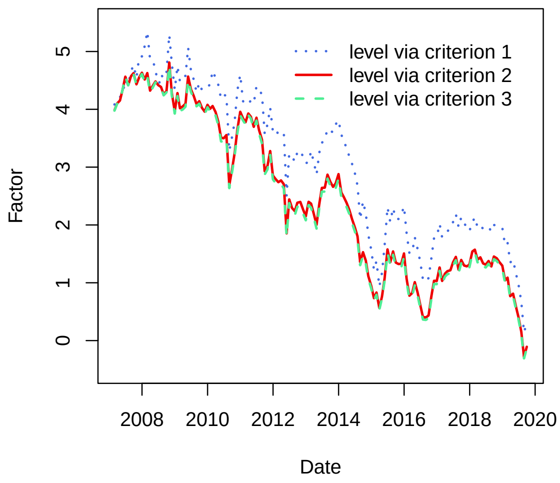


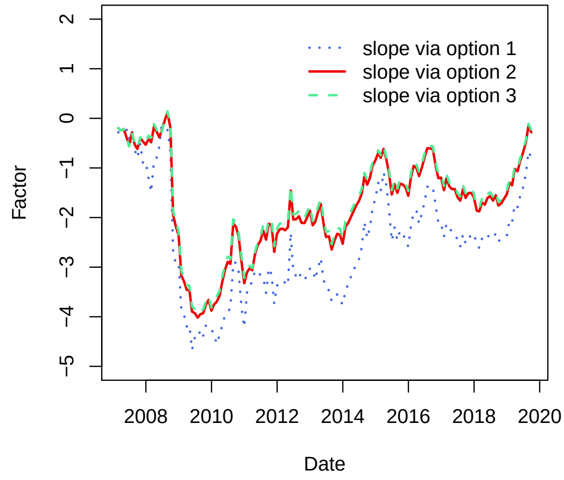
Figure 8: Average training set yield curve. Sample from March 2009 to September 2019.

To have an idea of the differences between the criteria, figure 9a, 9b, and 9c show the differences between the estimated factors of the initial training set yield curve following the three criteria.

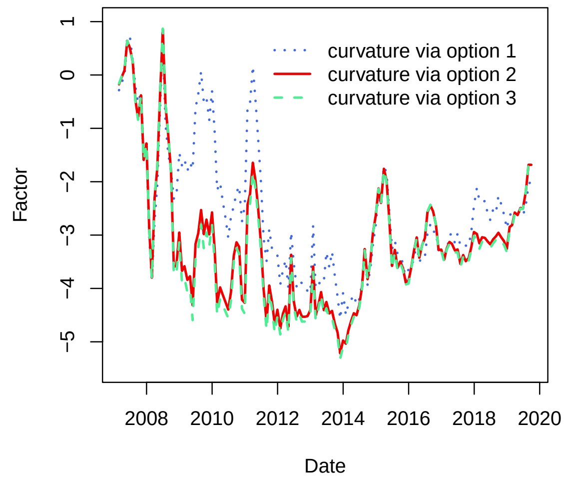
[9] If we were to use yearly maturities instead of monthly, we would have calculated $\lambda^* = 0.7173$.



(a) Level factor



(b) Slope factor



(c) Curvature factor

Figure 9: Estimated factors of the initial training set yield curve via the three criteria in the context of two-step estimation. *Note*: For criterion 1 we use here the MAE as the error statistic. In addition, we work with percentage yields, if we were to use yields expresses in regular units we would simply reduce all factors by a factor of 100, however the choice is inconsequential when it comes to forecasting applications.

In this case, it is not a surprise that the factors derived from criteria 2 and 3 are remarkably close: the optimal λ for criteria 2 and 3 is almost the same. However, if one changed the maturity of maximum curvature in criterion 3, then the factors would start to diverge more and more from the ones derived with criterion 2. Overall, all factors derived from the three criteria follow a similar path and present similar variations. However, the amplitude of those variations is different. For instance, the curvature factors derived from criterion 1 almost exactly match the curvature factors derived from the other two criteria at first, then diverge considerably, but then afterwards converge again (see figure 9c).

We can deduce as well some financial implications from the comparison of the factors derived from the three criteria. From a financial perspective, the level factors derived from criterion 1 suggest that the long term interest rates are higher, oftentimes about 1% higher (see figure 9a). In addition, the slope factors derived from criterion 1 suggest that the yield curve is generally more upward sloping (we deduce this because the slope factor is generally more negative). As an interesting additional commentary, we see that the sign of the slope factor becomes closer and closer to being positive at the times of macroeconomic turmoil (eg. 2008 and 2020), which means that the yield curve was near downward sloping. This is the well known yield curve inversion phenomenon.

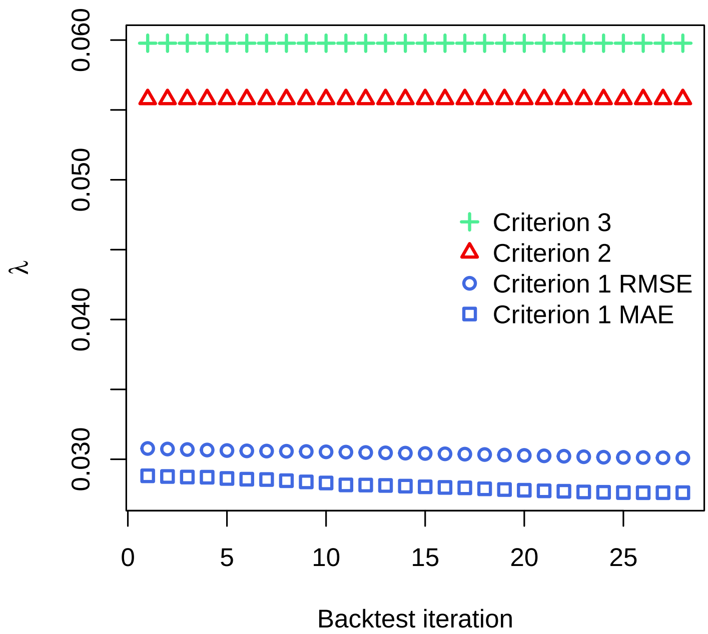


Figure 10: Optimal shape parameter λ^* for each backtest iteration following the three criteria. *Note*: As explained earlier, the values of λ^* for criteria 2 and 3 are constant through the backtest.

From the forecasting perspective, the differences in λ do change the factors' time series, which could have an impact on the coefficients from the autoregressive models. The following are the out-of-sample errors of the two-step IDNS and CDNS backtests following the three criteria:

1-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.020	0.014	0.028	0.059	0.073	0.077	0.080	0.084	0.087	0.090	0.093	0.096	0.098	0.101	0.110
2S IDNS-1$_{RMSE}$	0.022	0.014	0.031	0.064	0.076	0.079	0.080	0.084	0.087	0.090	0.093	0.096	0.099	0.101	0.113
2S IDNS-2	0.037	0.019	0.074	0.105	0.090	0.076	0.079	0.087	0.091	0.093	0.093	0.097	0.103	0.115	0.158
2S IDNS-3	0.039	0.021	0.082	0.110	0.090	0.076	0.080	0.087	0.092	0.093	0.093	0.097	0.104	0.119	0.165
2S CDNS-1$_{MAE}$	0.044	0.031	0.024	0.052	0.067	0.074	0.078	0.084	0.088	0.092	0.095	0.097	0.098	0.100	0.105
2S CDNS-1$_{RMSE}$	0.049	0.033	0.025	0.053	0.069	0.074	0.078	0.084	0.089	0.092	0.096	0.097	0.098	0.100	0.106
2S CDNS-2	0.090	0.038	0.039	0.074	0.072	0.072	0.086	0.098	0.103	0.102	0.100	0.097	0.099	0.104	0.137
2S CDNS-3	0.094	0.037	0.042	0.076	0.071	0.073	0.089	0.101	0.105	0.104	0.100	0.097	0.099	0.105	0.143

3-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.025	0.024	0.047	0.109	0.138	0.148	0.151	0.156	0.162	0.169	0.176	0.182	0.186	0.189	0.196
2S IDNS-1$_{RMSE}$	0.025	0.024	0.052	0.117	0.145	0.152	0.154	0.156	0.161	0.168	0.175	0.181	0.186	0.189	0.198
2S IDNS-2	0.027	0.041	0.120	0.186	0.175	0.153	0.143	0.148	0.156	0.163	0.169	0.175	0.185	0.195	0.226
2S IDNS-3	0.023	0.048	0.131	0.193	0.176	0.148	0.142	0.145	0.152	0.159	0.167	0.175	0.186	0.196	0.231
2S CDNS-1$_{MAE}$	0.090	0.070	0.053	0.090	0.120	0.135	0.146	0.157	0.167	0.176	0.184	0.191	0.195	0.198	0.203
2S CDNS-1$_{RMSE}$	0.098	0.074	0.052	0.090	0.120	0.135	0.146	0.157	0.168	0.177	0.185	0.191	0.195	0.197	0.201
2S CDNS-2	0.161	0.096	0.047	0.099	0.119	0.137	0.157	0.173	0.183	0.189	0.191	0.191	0.191	0.191	0.195
2S CDNS-3	0.168	0.097	0.048	0.100	0.119	0.139	0.160	0.175	0.184	0.189	0.191	0.190	0.190	0.191	0.195

6-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.049	0.074	0.147	0.267	0.316	0.332	0.335	0.338	0.346	0.353	0.361	0.369	0.377	0.384	0.397
2S IDNS-1$_{RMSE}$	0.047	0.076	0.153	0.281	0.327	0.341	0.342	0.341	0.346	0.354	0.361	0.369	0.377	0.384	0.396
2S IDNS-2	0.055	0.119	0.259	0.389	0.393	0.363	0.334	0.320	0.326	0.335	0.345	0.355	0.367	0.380	0.417
2S IDNS-3	0.044	0.129	0.274	0.399	0.397	0.364	0.343	0.335	0.339	0.347	0.358	0.370	0.383	0.398	0.433
2S CDNS-1$_{MAE}$	0.142	0.137	0.166	0.250	0.292	0.312	0.322	0.332	0.343	0.354	0.365	0.373	0.382	0.389	0.403
2S CDNS-1$_{RMSE}$	0.151	0.142	0.166	0.250	0.292	0.312	0.322	0.332	0.343	0.355	0.365	0.373	0.382	0.388	0.400
2S CDNS-2	0.229	0.170	0.158	0.248	0.291	0.311	0.326	0.341	0.351	0.360	0.366	0.370	0.375	0.377	0.383
2S CDNS-3	0.238	0.171	0.157	0.249	0.291	0.311	0.328	0.342	0.351	0.360	0.366	0.369	0.374	0.375	0.383

12-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.443	0.552	0.718	0.875	0.912	0.915	0.906	0.895	0.883	0.875	0.871	0.868	0.866	0.862	0.836
2S IDNS-1$_{RMSE}$	0.442	0.558	0.734	0.896	0.932	0.931	0.920	0.907	0.894	0.885	0.881	0.879	0.878	0.875	0.852
2S IDNS-2	0.474	0.651	0.893	1.061	1.048	0.999	0.952	0.918	0.900	0.894	0.898	0.905	0.915	0.923	0.933
2S IDNS-3	0.478	0.669	0.917	1.079	1.057	1.003	0.953	0.920	0.904	0.900	0.905	0.915	0.926	0.935	0.948
2S CDNS-1$_{MAE}$	0.544	0.633	0.746	0.852	0.873	0.872	0.865	0.854	0.842	0.833	0.827	0.823	0.818	0.811	0.778
2S CDNS-1$_{RMSE}$	0.550	0.636	0.747	0.852	0.873	0.872	0.864	0.853	0.841	0.831	0.826	0.822	0.817	0.811	0.780
2S CDNS-2	0.602	0.655	0.745	0.851	0.870	0.865	0.855	0.844	0.832	0.823	0.818	0.816	0.815	0.813	0.805
2S CDNS-3	0.607	0.655	0.744	0.851	0.869	0.864	0.854	0.843	0.832	0.822	0.817	0.815	0.815	0.815	0.810

Table 3: Out-of-sample MAE of the forecasts produced by the two-step IDNS and CDNS following the three criteria for $h = 1, 3, 6, 12$ months. We highlight the smallest value for each maturity. 2S IDNS-1$_{\mathrm{MAE}}$ denotes the IDNS using MAE as the minimized error statistic for criterion 1, 2S IDNS-1$_{\mathrm{RMSE}}$ denotes the IDNS using RMSE as the minimized error statistic for criterion 1, 2S IDNS-2 denotes the IDNS following criterion 2 and 2S IDNS-3 denotes the IDNS following criterion 3. The same logic applies to 2S CDNS-1$_{\mathrm{MAE}}$, 2S CDNS-1$_{\mathrm{RMSE}}$, 2S CDNS-2 and 2S CDNS-3.

1-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.025	0.020	0.034	0.077	0.097	0.102	0.102	0.105	0.109	0.112	0.116	0.119	0.123	0.126	0.137
2S IDNS-1$_{RMSE}$	0.026	0.019	0.036	0.081	0.100	0.103	0.103	0.105	0.108	0.112	0.116	0.119	0.123	0.127	0.140
2S IDNS-2	0.042	0.024	0.079	0.120	0.114	0.100	0.099	0.107	0.113	0.115	0.117	0.120	0.127	0.138	0.184
2S IDNS-3	0.043	0.026	0.087	0.125	0.113	0.098	0.100	0.108	0.114	0.115	0.116	0.120	0.128	0.140	0.190
2S CDNS-1$_{MAE}$	0.047	0.036	0.034	0.070	0.090	0.097	0.101	0.106	0.112	0.117	0.121	0.123	0.125	0.126	0.132
2S CDNS-1$_{RMSE}$	0.052	0.038	0.034	0.071	0.091	0.097	0.101	0.107	0.113	0.118	0.121	0.123	0.125	0.126	0.133
2S CDNS-2	0.092	0.044	0.045	0.092	0.095	0.095	0.107	0.121	0.129	0.130	0.127	0.124	0.124	0.127	0.159
2S CDNS-3	0.096	0.043	0.049	0.094	0.094	0.095	0.110	0.124	0.131	0.131	0.127	0.123	0.124	0.128	0.164

3-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.033	0.031	0.072	0.189	0.236	0.246	0.244	0.243	0.245	0.249	0.254	0.260	0.265	0.267	0.273
2S IDNS-1$_{RMSE}$	0.033	0.031	0.075	0.194	0.240	0.249	0.245	0.243	0.245	0.248	0.253	0.259	0.265	0.267	0.275
2S IDNS-2	0.036	0.050	0.132	0.240	0.260	0.248	0.235	0.232	0.234	0.239	0.246	0.255	0.265	0.274	0.305
2S IDNS-3	0.029	0.053	0.141	0.244	0.258	0.244	0.231	0.228	0.230	0.236	0.243	0.253	0.264	0.274	0.309
2S CDNS-1$_{MAE}$	0.095	0.077	0.071	0.166	0.211	0.222	0.222	0.225	0.229	0.236	0.243	0.249	0.254	0.256	0.260
2S CDNS-1$_{RMSE}$	0.102	0.081	0.071	0.167	0.211	0.222	0.222	0.224	0.230	0.236	0.243	0.249	0.254	0.255	0.259
2S CDNS-2	0.165	0.103	0.065	0.175	0.211	0.217	0.221	0.230	0.237	0.242	0.246	0.248	0.250	0.251	0.261
2S CDNS-3	0.171	0.104	0.065	0.176	0.210	0.217	0.222	0.231	0.239	0.243	0.246	0.247	0.249	0.250	0.263

6-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.076	0.158	0.311	0.505	0.568	0.586	0.588	0.587	0.587	0.590	0.595	0.602	0.609	0.613	0.614
2S IDNS-1$_{RMSE}$	0.075	0.159	0.315	0.512	0.574	0.591	0.591	0.589	0.588	0.590	0.596	0.602	0.610	0.614	0.617
2S IDNS-2	0.085	0.190	0.373	0.569	0.606	0.599	0.584	0.575	0.573	0.577	0.587	0.599	0.613	0.625	0.649
2S IDNS-3	0.077	0.192	0.381	0.575	0.608	0.597	0.582	0.573	0.572	0.577	0.587	0.600	0.615	0.628	0.655
2S CDNS-1$_{MAE}$	0.150	0.172	0.285	0.464	0.523	0.539	0.541	0.540	0.540	0.543	0.549	0.555	0.562	0.565	0.566
2S CDNS-1$_{RMSE}$	0.159	0.174	0.284	0.464	0.523	0.539	0.540	0.539	0.539	0.542	0.548	0.554	0.561	0.565	0.566
2S CDNS-2	0.238	0.193	0.281	0.472	0.521	0.528	0.526	0.526	0.529	0.535	0.542	0.550	0.560	0.566	0.577
2S CDNS-3	0.246	0.193	0.281	0.472	0.520	0.526	0.524	0.525	0.528	0.534	0.542	0.550	0.560	0.567	0.579

12-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS-1$_{MAE}$	0.926	1.112	1.320	1.427	1.414	1.386	1.356	1.325	1.296	1.273	1.256	1.244	1.233	1.220	1.165
2S IDNS-1$_{RMSE}$	0.927	1.116	1.330	1.441	1.429	1.399	1.367	1.334	1.303	1.280	1.263	1.252	1.242	1.229	1.178
2S IDNS-2	0.953	1.167	1.414	1.537	1.498	1.438	1.384	1.340	1.307	1.286	1.278	1.276	1.278	1.277	1.261
2S IDNS-3	0.954	1.177	1.434	1.560	1.517	1.454	1.399	1.355	1.324	1.305	1.298	1.298	1.301	1.302	1.289
2S CDNS-1$_{MAE}$	0.895	1.074	1.278	1.381	1.366	1.336	1.303	1.270	1.238	1.212	1.193	1.179	1.166	1.151	1.091
2S CDNS-1$_{RMSE}$	0.892	1.071	1.277	1.383	1.367	1.336	1.301	1.267	1.234	1.209	1.190	1.177	1.165	1.151	1.096
2S CDNS-2	0.870	1.050	1.277	1.397	1.365	1.315	1.270	1.232	1.203	1.184	1.175	1.171	1.170	1.167	1.142
2S CDNS-3	0.868	1.049	1.279	1.398	1.363	1.310	1.264	1.227	1.199	1.182	1.174	1.172	1.172	1.170	1.149

Table 4: Out-of-sample RMSE of the forecasts produced by the two-step IDNS and CDNS following the three criteria for $h = 1, 3, 6, 12$ months. We highlight the smallest value for each maturity. 2S IDNS-1$_{\mathrm{MAE}}$ denotes the IDNS using MAE as the minimized error statistic for criterion 1, 2S IDNS-1$_{\mathrm{RMSE}}$ denotes the IDNS using RMSE as the minimized error statistic for criterion 1, 2S IDNS-2 denotes the IDNS following criterion 2 and 2S IDNS-3 denotes the IDNS following criterion 3. The same logic applies to 2S CDNS-1$_{\mathrm{MAE}}$, 2S CDNS-1$_{\mathrm{RMSE}}$, 2S CDNS-2 and 2S CDNS-3.

We can see that each criteria allows for different forecast accuracy across maturities and across forecast horizons. For the IDNS, criterion 1 with minimized MAE seems to be superior for most maturities, specially at $h = 12$. For the CDNS, criterion 1 with MAE minimization also seems to be superior criterion for most maturities, although criterion 3 is superior for an important number of maturities, specially at $h = 12$.

	h = 1	h = 3	h = 6	h = 12	Average
2S IDNS-1_{MAE}	0.074	0.137	0.296	0.819	0.332
2S IDNS-1_{RMSE}	0.075	0.139	0.300	0.831	0.336
2S IDNS-2	0.088	0.151	0.317	0.891	0.362
2S IDNS-3	0.090	0.151	0.327	0.901	0.367
2S CDNS-1_{MAE}	0.075	0.145	0.304	0.798	0.331
2S CDNS-1_{RMSE}	0.076	0.146	0.305	0.798	0.331
2S CDNS-2	0.087	0.155	0.310	0.801	0.338
2S CDNS-3	0.089	0.156	0.311	0.801	0.339

Table 5: Average out-of-sample MAE across all maturities for the two-step IDNS and CDNS following the three criteria. The *Average* column calculates the average across $h = 1, 3, 6, 12$.

	h = 1	h = 3	h = 6	h = 12	Average
2S IDNS-1_{MAE}	0.094	0.207	0.506	1.264	0.518
2S IDNS-1_{RMSE}	0.095	0.208	0.508	1.273	0.521
2S IDNS-2	0.107	0.217	0.520	1.313	0.539
2S IDNS-3	0.108	0.216	0.521	1.331	0.544
2S CDNS-1_{MAE}	0.096	0.201	0.473	1.209	0.495
2S CDNS-1_{RMSE}	0.097	0.202	0.473	1.208	0.495
2S CDNS-2	0.107	0.208	0.476	1.199	0.498
2S CDNS-3	0.109	0.209	0.476	1.198	0.498

Table 6: Average out-of-sample RMSE across all maturities for the two-step IDNS and CDNS following the three criteria. The *Average* column calculates the average across $h = 1, 3, 6, 12$.

For both IDNS and CDNS, criterion 1 with the MAE as the minimized error statistic had the best forecasting performance out of all criteria: the IDNS following criterion 1 (for both minimized MAE and RMSE) had around 10% lower out-of-sample MAE than criterion 2 and 3, while for the CDNS it was around 2% lower. In terms of out-of-sample RMSE, the criterion 1 with minimized MAE obtained the best results, with the exception of criterion 3 for the CDNS at 12-month-ahead horizon. Therefore, we consider criterion 1 with MAE minimization as the winner of this competition.

Let's analyse in detail the two-step DNS following criteria 1 using the MAE as the minimized error statistic.

Factor	ADF
l	0.063
s	0.215
c	0.298

Table 7: Augmented Dickey-Fulley test p-values for the series of factors obtained via the two-step approach for the initial training set yield curve. The alternative hypothesis is the stationarity of the series. Recall that both the two-step IDNS and CDNS use the same time series of factors.

The ADF test does not suggest the stationarity of the factors' time series based on a confidence level of 95%: there may be a unit root on each serie. Note that this conclusion may change as the backtest progresses and new data is integrated into the training set. An examination of the ACF and PACF function (see figures 24 and 25 on appendix E) suggests that an AR type model could be more appropriate than a MA type model due to the progressive decay of the ACF function in conjunction with the sharp decay of the PACF function (see Shumway and Stoffer, 2016).

	l	s	c
l	1.000		
s	-0.296	1.000	
c	0.396	0.255	1.000

Table 8: Correlation matrix of the factors estimated through OLS on the initial training set yield curve. Recall that these are the factors' time series used by the two-step IDNS and CDNS.

It is worth noting that there is considerable correlation between the factors, as it can be seen in table 8. Nonetheless, this doesn't mean that the CDNS would be more appropriate than the IDNS because table 8 displays the correlation between present factors, while the CDNS needs high correlation between present values of a specific factor and past values of the other factors.

Level factor

Model	AIC	BIC
AR(1)	45.1	54.1
AR(2)	45.7	57.8
MA(1)	346.5	355.6

Slope factor

Model	AIC	BIC
AR(1)	101.0	110.1
AR(2)	102.7	114.8
MA(1)	315.6	324.6

Curvature factor

Model	AIC	BIC
AR(1)	236.8	245.9
AR(2)	237.9	250.0
MA(1)	373.0	382.0

All factors

Model	AIC	BIC
VAR(1)	-7.8	-7.6
VAR(2)	-8.2	-7.8
VAR(3)	-8.1	-7.5

Table 9: Akaike Information Criterion and Bayesian Information Criterion of some autoregressive models for the individual factors and for all factors. We highlight the lowest value for each criterion.

The results from table 9 suggest that the AR(1) model is the most adequate autoregressive model for the IDNS based on the minimization of the AIC and BIC criteria. For the CDNS, the results suggest that a VAR(2) model would be more adequate.

We perform portmanteau tests on the residuals of the AR(1) for each factor as well as on the residuals of the VAR(1) for all factors (see figures 26 and 27 on appendix E for plots of the residuals).

AR(1)	
Factor	Portmanteau test
l	0.372
s	0.289
c	0.011

VAR(1)	
Factor	Portmanteau test
All factors	4.063e-14

Table 10: Portmanteau test p-value of the residuals of the AR(1) and VAR(1) fits to the initial training set yield curve. The alternative hypothesis is the non-null autocorrelation of the residuals. *Note*: for the AR(1) residuals we use the Ljung-Box test with 10 lags and 9 degrees of freedom. For the VAR(1) residuals we use a multivariate portmanteau test with 10 lags.

Based on table 10, only the residuals of the curvature factor have statistically significant autocorrelation. This suggests that at least for the level and slope factors we have some evidence that the AR(1) captures adequately the temporal dependence of the data. On the other hand, the VAR(1) residuals do have statistically significant autocorrelation. In addition, a visual inspection on the residuals (appendix E, figures 26 and 27) hints the possible heteroscedasticity of the residuals, specially for the curvature factor residuals.

Overall, we observe some indications that the chosen models might not be adequate. Noticeably, no differentiation is applied to the series even though we have no conclusive evidence of the stationarity of the factors. In addition, the AIC and BIC minimization criteria suggest that a VAR of first order is less suitable than a VAR of second order, not to mention that the multivariate portmanteau test suggests that the VAR(1) residuals are autocorrelated. Despite of this, we also have some evidence of the adequacy of the models: the AIC and BIC minimization criteria does support a first order for the AR model and the residuals of both models seem to be normally distributed.

4.2 Two-step vs. one-step: empirical assessment

4.2.1 Forecasting performance based on market data

Given the superiority of criterion 1, we compare the out-of-sample error of the two-step IDNS and CDNS following criterion 1 with MAE minimization against the one-step IDNS and CDNS. As mentioned earlier, we also include the Random Walk as a general benchmark.

1-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.022	0.014	0.031	0.064	0.076	0.079	0.080	0.084	0.087	0.090	0.093	0.096	0.099	0.101	0.113
2S CDNS	0.049	0.033	0.025	0.053	0.069	0.074	0.078	0.084	0.089	0.092	0.096	0.097	0.098	0.100	0.106
1S IDNS	0.042	0.028	0.024	0.059	0.077	0.083	0.084	0.086	0.089	0.091	0.094	0.097	0.099	0.101	0.108
1S CDNS	0.044	0.034	0.035	0.064	0.078	0.085	0.087	0.090	0.094	0.097	0.099	0.101	0.103	0.105	0.114
Random Walk	0.008	0.012	0.024	0.051	0.064	0.073	0.080	0.086	0.090	0.094	0.097	0.099	0.100	0.101	0.105

3-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.025	0.024	0.052	0.117	0.145	0.152	0.154	0.156	0.161	0.168	0.175	0.181	0.186	0.189	0.198
2S CDNS	0.098	0.074	0.052	0.090	0.120	0.135	0.146	0.157	0.168	0.177	0.185	0.191	0.195	0.197	0.201
1S IDNS	0.047	0.035	0.045	0.110	0.143	0.154	0.158	0.161	0.164	0.169	0.174	0.180	0.185	0.187	0.193
1S CDNS	0.092	0.081	0.080	0.126	0.151	0.162	0.166	0.170	0.175	0.180	0.185	0.190	0.193	0.195	0.198
Random Walk	0.009	0.016	0.037	0.088	0.115	0.133	0.143	0.153	0.161	0.167	0.174	0.179	0.183	0.186	0.191

6-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.047	0.076	0.153	0.281	0.327	0.341	0.342	0.341	0.346	0.354	0.361	0.369	0.377	0.384	0.396
2S CDNS	0.151	0.142	0.166	0.250	0.292	0.312	0.322	0.332	0.343	0.355	0.365	0.373	0.382	0.388	0.400
1S IDNS	0.069	0.085	0.143	0.265	0.317	0.334	0.337	0.340	0.344	0.350	0.357	0.365	0.373	0.379	0.391
1S CDNS	0.159	0.165	0.216	0.317	0.356	0.370	0.375	0.377	0.380	0.384	0.390	0.396	0.403	0.409	0.421
Random Walk	0.026	0.061	0.131	0.241	0.289	0.315	0.330	0.341	0.350	0.358	0.366	0.372	0.379	0.383	0.393

12-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.442	0.558	0.734	0.896	0.932	0.931	0.920	0.907	0.894	0.885	0.881	0.879	0.878	0.875	0.852
2S CDNS	0.550	0.636	0.747	0.852	0.873	0.872	0.864	0.853	0.841	0.831	0.826	0.822	0.817	0.811	0.780
1S IDNS	0.462	0.565	0.713	0.872	0.912	0.916	0.908	0.897	0.887	0.879	0.875	0.872	0.870	0.865	0.837
1S CDNS	0.595	0.699	0.839	0.965	0.989	0.988	0.980	0.969	0.958	0.948	0.942	0.936	0.931	0.923	0.888
Random Walk	0.436	0.554	0.715	0.858	0.897	0.913	0.922	0.926	0.928	0.931	0.934	0.937	0.939	0.938	0.921

Table 11: Out-of-sample MAE of the forecasts. In gray the lowest MAE by maturity. Note: 2S IDNS and 2S CDNS are the two-step IDNS and CDNS while the 1S IDNS and 1S CDNS are the one-step IDNS and CDNS. Also note that we work with percentage yields, meaning that these MAE reflect the difference in the percentage.

1-month-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.026	0.019	0.036	0.081	0.100	0.103	0.103	0.105	0.108	0.112	0.116	0.119	0.123	0.127	0.140
2S CDNS	0.052	0.038	0.034	0.071	0.091	0.097	0.101	0.107	0.113	0.118	0.121	0.123	0.125	0.126	0.133
1S IDNS	0.048	0.034	0.032	0.078	0.103	0.108	0.108	0.109	0.111	0.114	0.117	0.120	0.124	0.126	0.134
1S CDNS	0.056	0.046	0.046	0.081	0.101	0.105	0.106	0.110	0.114	0.119	0.123	0.126	0.129	0.130	0.137
Random Walk	0.013	0.018	0.031	0.070	0.085	0.094	0.101	0.107	0.112	0.116	0.119	0.121	0.124	0.125	0.131

3-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.033	0.031	0.075	0.194	0.240	0.249	0.245	0.243	0.245	0.248	0.253	0.259	0.265	0.267	0.275
2S CDNS	0.102	0.081	0.071	0.167	0.211	0.222	0.222	0.224	0.230	0.236	0.243	0.249	0.254	0.255	0.259
1S IDNS	0.055	0.043	0.069	0.189	0.239	0.250	0.247	0.245	0.245	0.247	0.251	0.256	0.261	0.262	0.266
1S CDNS	0.126	0.115	0.120	0.201	0.239	0.246	0.242	0.240	0.241	0.245	0.249	0.253	0.257	0.258	0.260
Random Walk	0.015	0.025	0.069	0.169	0.207	0.225	0.231	0.238	0.244	0.249	0.253	0.257	0.261	0.261	0.265

6-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.075	0.159	0.315	0.512	0.574	0.591	0.591	0.589	0.588	0.590	0.596	0.602	0.610	0.614	0.617
2S CDNS	0.159	0.174	0.284	0.464	0.523	0.539	0.540	0.539	0.539	0.542	0.548	0.554	0.561	0.565	0.566
1S IDNS	0.089	0.158	0.307	0.503	0.567	0.585	0.587	0.585	0.583	0.584	0.588	0.594	0.600	0.603	0.602
1S CDNS	0.221	0.249	0.348	0.511	0.563	0.576	0.574	0.571	0.569	0.570	0.573	0.578	0.584	0.586	0.585
Random Walk	0.067	0.157	0.308	0.478	0.532	0.558	0.571	0.580	0.586	0.592	0.598	0.603	0.610	0.613	0.614

12-months-ahead

	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	11Y	12Y	15Y
2S IDNS	0.927	1.116	1.330	1.441	1.429	1.399	1.367	1.334	1.303	1.280	1.263	1.252	1.242	1.229	1.178
2S CDNS	0.892	1.071	1.277	1.383	1.367	1.336	1.301	1.267	1.234	1.209	1.190	1.177	1.165	1.151	1.096
1S IDNS	0.926	1.110	1.318	1.423	1.411	1.383	1.353	1.322	1.293	1.269	1.253	1.240	1.229	1.216	1.160
1S CDNS	0.962	1.137	1.338	1.440	1.422	1.389	1.354	1.319	1.286	1.259	1.239	1.224	1.210	1.195	1.135
Random Walk	0.945	1.132	1.334	1.420	1.403	1.389	1.376	1.361	1.345	1.331	1.320	1.312	1.305	1.294	1.246

Table 12: Out-of-sample RMSE of the forecasts. In gray the lowest RMSE by maturity. Note: 2S IDNS and 2S CDNS are the two-step IDNS and CDNS while the 1S IDNS and 1S CDNS are the one-step IDNS and CDNS. Also note that we work with percentage yields, meaning that these RMSE reflect the difference in the percentage.

We observe that there is a generalized increase in error as the horizon increases (see appendix 8.7 for a visual examination of the fits). It was reasonable to expect this since forecasts usually deteriorate as they predict further and further into the future. Papers like Diebold and Li, 2006 and Caldeira and al., 2016 do the same empirical observation. We also observe that for each horizon, the short end of the yield curve is better forecasted than the medium and long end of the yield curve, these results also coincide with the results obtained by Diebold and Li, 2006 and Caldeira and al., 2016.

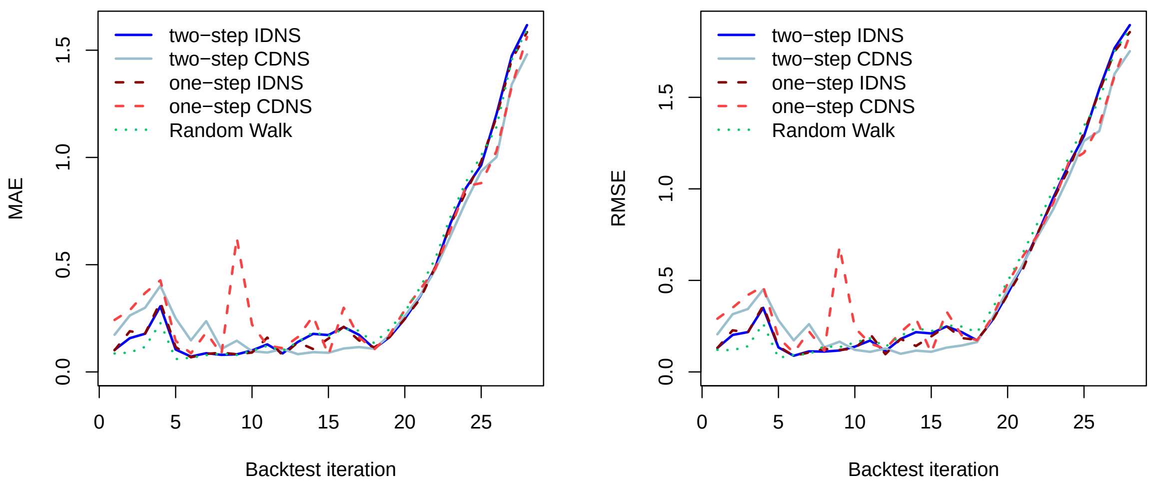
	h = 1	h = 3	h = 6	h = 12	Average
2S IDNS	0.075	0.139	0.300	0.831	0.336
2S CDNS	0.076	0.146	0.305	0.798	0.331
1S IDNS	0.077	0.140	0.297	0.822	0.334
1S CDNS	0.082	0.156	0.341	0.903	0.371
Random Walk	0.072	0.129	0.289	0.850	0.335

Table 13: Average out-of-sample MAE across all maturities of table 11. The *Average* column calculates the average across h = 1, 3, 6, 12.

	h = 1	h = 3	h = 6	h = 12	Average
2S IDNS	0.095	0.208	0.508	1.273	0.521
2S CDNS	0.097	0.202	0.473	1.208	0.495
1S IDNS	0.098	0.208	0.502	1.261	0.517
1S CDNS	0.102	0.220	0.510	1.261	0.523
Random Walk	0.091	0.198	0.498	1.301	0.522

Table 14: Average out-of-sample RMSE across all maturities of table 12. The *Average* column calculates the average across h = 1, 3, 6, 12.

In terms of the forecasting performance of each model, the one-step CDNS produces less accurate results than the one-step IDNS, its two-step counterparts and the Random Walk. Nonetheless, despite also relying on the same optimization algorithm as the one-step CDNS, the one-step IDNS provides slightly better results than the two-step IDNS. In general, the two-step CDNS provides the best results among all models for forecast horizons of 6 and 12 months ahead. These results are in line with previous results from Caldeira and al., 2016 and Christensen, J.H.E., Diebold, X. and Rudebusch. G.D, 2011, who also empirically found that the CDNS can surpass the IDNS in forecasting performance. The Random Walk has a notably good forecasting performance with respect to every other model, notably for the short end of the yield curve. As maturity increases, the Random Walk remains nonetheless competitive.

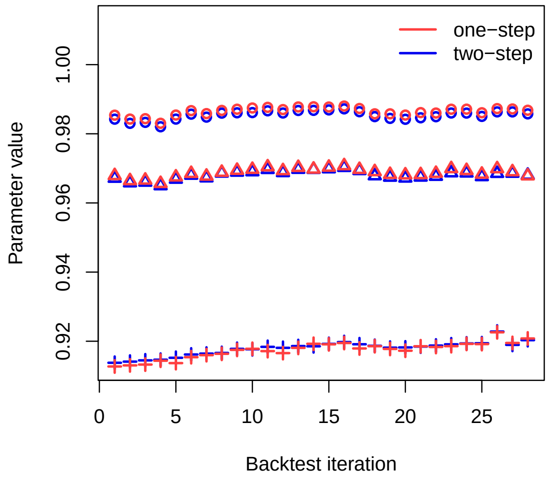


(a) Out-of-sample MAE

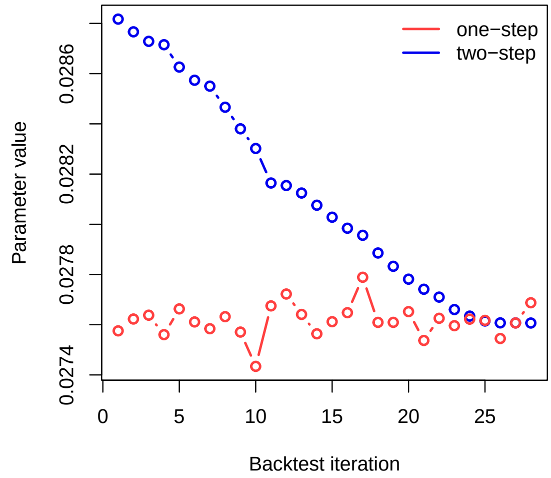
(b) Out-of-sample RMSE

Figure 11: Out-of-sample error on each backtest iteration. We condense for each iteration all the errors of every horizon from 1 month up to 12 months.

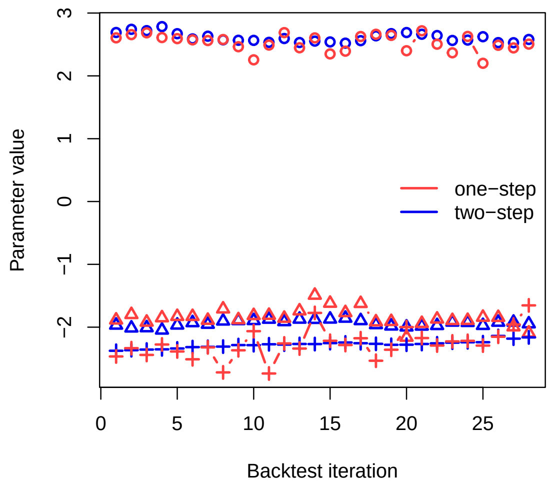
One can see in figure 11 that most of the errors occur when the yield curve increases around 2022, during this steady increase all models perform essentially the same. Considering the complete backtest, the one-step CDNS seems to produce more sporadic errors, notably at iteration 9. Interestingly, the errors produced by the two-step and one-step IDND are very similar. Contrariwise, the errors produced by the two-step and one-step CDNS are considerably different. Both IDNS models and the Random walk produce similar errors in terms of magnitude and stability. Judging by the results from tables 11 and 12 and figure 11, we are of the opinion that the superiority of one model with respect to another should not only be decided based on the error incurred in a particular maturity and forecast horizon, but one should also consider the stability of the forecasts through the backtest.



(a) Circles: a_{11}, triangles: a_{22}, crosses: a_{33}



(b) λ



(c) Circles: μ^l, triangles: μ^s, crosses: μ^c

Figure 12: Values of A, λ and μ for the two-step IDNS and one-step IDNS at each backtest iteration. *Note*: in the one-step appraoch we estimate every parameter at each backtest iteration. In the two-step approach some parameters such as λ, A, μ are estimated and used for the forecasting, while other parameters such as Q and H are implied by the data and the fit of the model, but are not used to produce forecasts.

Figures 12a, 12b and 12c provide more insight about the way the two-step IDNS and one-step IDNS work (see figures 28a, 28b, 28c and 28d in appendix 8.6 for the rest of the parameters). Considering all parameters, it is easy to see that the parameters estimated via one-step approach are more unstable than the ones estimated or implied by the two-step approach. Despite of this, we already observed that the one-step IDNS achieved better forecasting performance than the two-step IDNS. Notwithstanding the good forecasting performance of the one-step IDNS, the one-step CDNS produced highly unstable errors. As discussed by Diebold and Rudebusch, 2013, the optimization algorithm may not have converged in a trustworthy way, producing highly volatile parameter estimates and therefore highly volatile forecasts and errors. One should be careful when using this kind of procedure as the estimations can be heavily incorrect on a sporadically fashion.

4.2.2 Parameter estimation based on simulated data

To give additional insight on how accurate both estimation methods are, we assess the ability of the IDNS following both the two-step and the one-step approaches to find the actual parameters of a number of simulated yield curves. The objective is to determine if the estimation methods allow for good performance under ideal conditions. Indeed, if the models were to produce accurate results, we could be more confident in their usefulness with market data. Contrariwise, if the models were to perform poorly in capturing the actual parameters, then even if the market factors followed the ideal underlying dynamics we could not be certain of the accuracy of the results.
In this controlled simulation exercise, we generate artificial yield curves by simulating the path of the three factors and then using the Nelson-Siegel model to fit a yield curve. We repeat this process in a Monte-Carlo fashion to produce a distribution of estimated parameter values for the IDNS following the two-step and the one-step approaches. In addition, we naturally expect both approaches to improve their estimations the more data is input into them. We therefore carry the previously mentioned exercise under four scenarios where we vary the number of observations each simulated yield curve has. In the first scenario we fix a window of 240 observations per yield curve (corresponding to 20 years of data), in the second scenario we fix a window of 120 observations (10 years of data), in the third one we fix a window of 60 observations (5 years of data) and in the fourth scenario we fix a window of 30 observations (2.5 years of data). In that way, we can assess how both the two-step and the one-step approach perform under ideal conditions given a certain length of the input data, as it may happen in any practical implementation.
Each one of our simulated factors follows an AR(1) process such as the one in equation

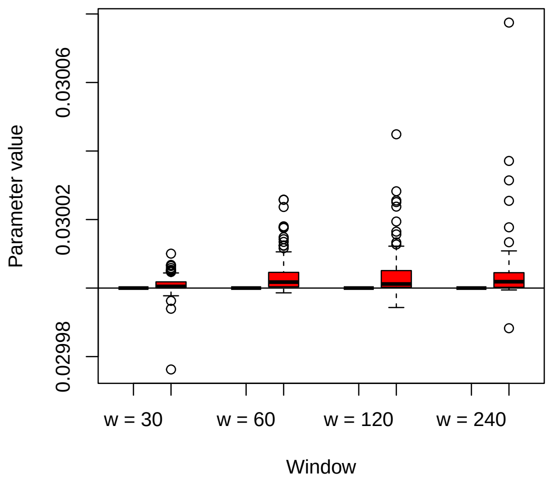
(14) with the following characteristics:

$$A = \begin{pmatrix} 0.95 & 0 & 0 \\ 0 & 0.95 & 0 \\ 0 & 0 & 0.95 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.25 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.25 \end{pmatrix}, \quad \mu = \begin{pmatrix} 4 \\ -2 \\ -3 \end{pmatrix}, \quad f_1 = \begin{pmatrix} 6 \\ -0.25 \\ -0.30 \end{pmatrix} \tag{33}$$

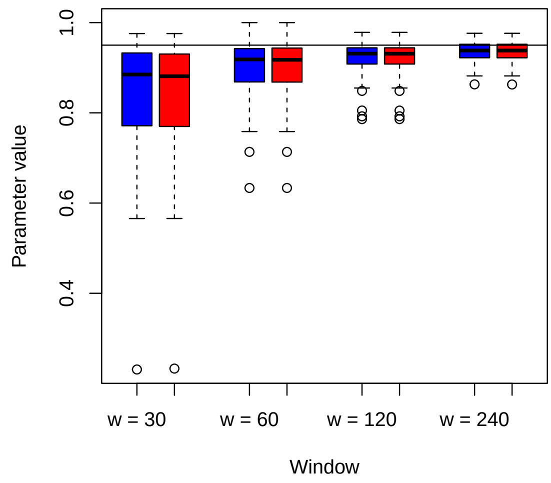
In other words, these are the true factors our models try to estimate. We simulate gaussian factor disturbances at every time $t > 1$ such that:

$$\eta_t \sim \mathcal{N}(0, Q) \tag{34}$$

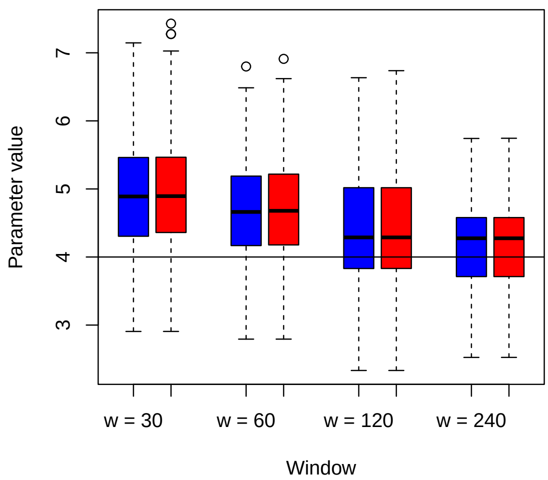
We generate 100 paths for each of the three factors, which we transform into 100 yield curves using equation (13). We then input a window of each yield curve into the two-step and one-step IDNS just as described and performed in our previous procedure. Note that the IDNS assumes that the factors follow precisely an AR(1) process and the yields are given by the Nelson-Siegel model. Our simulation is therefore a very favorable scenario for the IDNS to operate in. The results are the following:



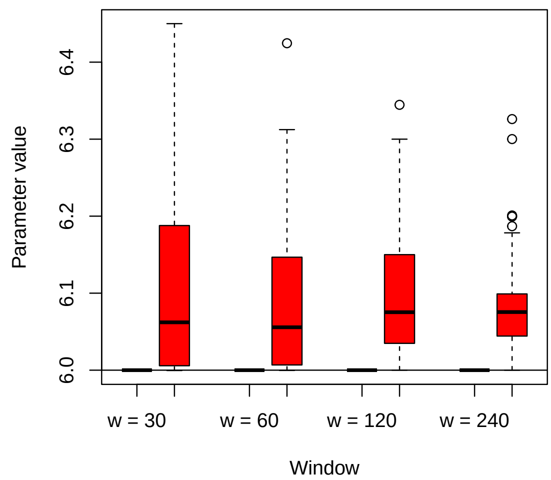
(a) λ



(b) a_{11}



(c) μ^l



(d) l_1

Figure 13: Estimated parameters λ, a_{11}, μ^l and l_1 extracted from 100 simulated yield curves for window lengths of 30, 60, 120 and 240 observations. In blue: two-step IDNS, in red: one-step IDNS. The horizontal line is the value of the true parameter.

We display in figure 13 a couple of representative results of the simulation exercise (see appendix 8.8 for the rest of the estimated parameters). Usually, the bigger the window, meaning the more observation points the yield curve has, the more accurate the estimation of the parameters. In this particular example, the starting point of each factor was usually not the same as their long term mean, as a consequence the first observations were closer to that starting point, giving the impression that the long term mean and autoregressive coefficient were different than their actual value. That is the reason why

we observe some distributions of parameter values shift towards the actual parameter value as we increase the window length. In addition, the dispersion of the distribution of the estimated parameter values naturally decreases as the window increases because the model can use more and more data. Regarding the differences between the two-step and one-step estimation, some parameters such as μ^l in figure 13c and a_{11} in figure 13b are estimated with essentially the same accuracy via both approaches. However, some others such as λ in figure 13a and l_1 in figure 13d do present some differences: the one-step approach estimated the parameters with less accuracy than the two-step approach. In the case of λ, even though the dispersion of the estimated parameters via one-step approach is bigger than the one of the two-step, the magnitude of the differences is not dramatic. On the other hand, when it comes to l_1 the parameter was sometimes estimated as having almost 0.4 more units than what it really had, which is a considerable mismatch. It is possible that this phenomenon is due to the smoothing of the factors after the Kalman filter run with the optimal parameters. Indeed, the smoothing may change the value of the originally estimated parameters to better fit neighbour values, which tended to be closer to 0 due to the specifications of our AR(1) model in (33). In summary, even though the two-step and one-step approaches provide similar results, the one-step approach seems to produce more unstable parameter estimates than its two-step counterpart.
Despite being this a simulated exercise, one could also expect the one-step approach to deliver more unstable estimates in a real life scenario, which seems to be our case.

4.3 Robustness analysis results

4.3.1 Sensitivity of the two-step IDNS to λ

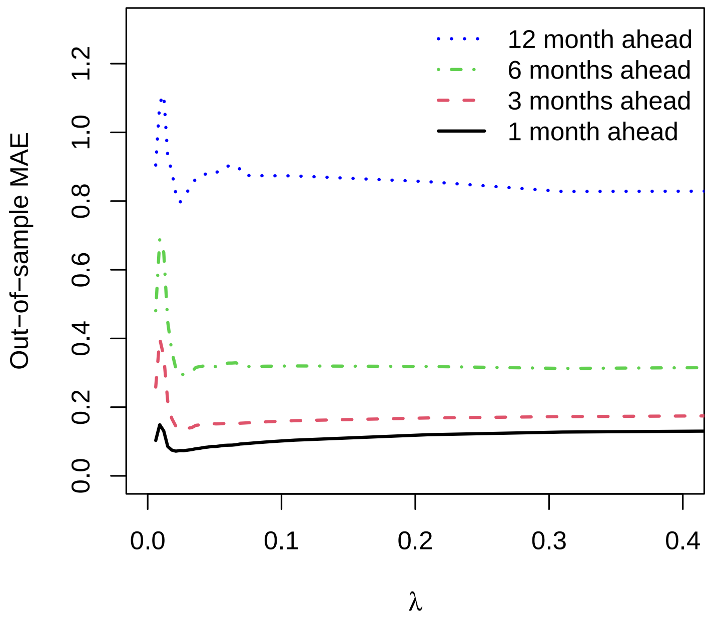


Figure 14: Out-of-sample MAE of the two-step IDNS model for $h = 1, 3, 6, 12$ as a function of the value of the fixed λ. *Note*: each MAE is calculated across an entire backtest, meaning that the the out-of-sample MAE value is calculated by fixing the value of λ for the entire backtest, then performing the backtest and after that calculating the MAE. This is different from the procedure explained in the forecast methodology section, in which we re-calibrate the value of λ at each step of the backtest.

Figure 14 displays the results of the robustness analysis with respect to λ. Visual examination suggests that the out-of-sample MAE is sensible to the value of λ, especially for small values. Note that for every horizon, a global minimum is visible next to the global maximum on the left end of the curve. Then, the MAE flattens as the value of λ increases. Naturally, the out-of-sample MAE increases as the forecast horizon increases as well, we observe this on the upward shifts each curve of higher horizon endures. For the curves corresponding to short forecast horizons, there is no much difference in MAE for the maximum and minimum of the curve: for the 1-month-ahead curve there is a difference in MAE of about 0.1 units. On the other hand, for the 6-months-ahead curve the difference reaches about 0.4 units. The difference is considerable: an improper value for λ could increase the forecasting error of the IDNS in such a way that the model becomes less accurate than all of the models considered in this work. To put it into perspective, the maximum-to-minimum MAE ratio of the 1-month-ahead curve is around 2 (meaning that sliding from the minimum to the maximum of the curve would increase the MAE in about 100%). The maximum-to-minimum MAE ratio of the 6-month-ahead curve is also around 2 (or an increase in MAE of about 100%). Careful inspection also reveals that the minima on each curve do not land on the same location, they are slightly shifted in the horizontal axis by an order of magnitude of about 0.01 units of λ. It is easy to see that an optimal value of λ for one horizon could be a not so optimal value for another horizon.

Given the sensibility of the forecast to the value of λ, we believe it's a good practice to plot the out-of-sample MAE of the two-step IDNS model for some of the forecasting horizons just as we did in figure 14. This allows the modeler to be aware if the optimal value of λ for one forecasting horizon is not the same as a non-optimal value for another forecasting horizon, which might substantially worsen its MAE by falling into the out-of-sample MAE global maximum. It is worth mentioning that even though the values of λ^* found via the three criteria discussed in this work are close to the optimal values seen in figure 14, there is no guarantee that they will be equal. Not to mention that some values of λ^* change at each backtest iteration because they are re-calibrated, such as the ones calculated via the in-sample error minimization criterion.

4.3.2 Sensitivity of the one-step IDNS to the initial parameters

As mentioned in the methodology, we apply shocks or deviations to every matrix (in the case of A, Q, H and P_1), vector (in the case of μ and f_1) and standalone parameter (in the case of λ) in the set of base parameters. The base parameters are the initialization parameters we obtained running the two-step IDNS as discussed in the methodology section. As it is well known, optimization algorithms might not reach convergence if the starting values of an optimization are too far away of the optimum. In addition, they might get stuck in local optima and produce different parameter estimates. Since the choice of the initial parameter guesses is subjective to a certain extent, it is of particular interest to determine how different are the results produced by the one-step IDNS depending on the starting value of the optimization. If the results were to be different, then we would conclude that the model produces unstable results and it should be viewed with more distrust. On the other hand, if the results were to be similar, then we would conclude that it is not primordial to spend that much time and effort in accurately estimating the already mentioned initial values. Bellow are the results:

	-50%	-20%	-10%	-5%	-1%	base	+1%	+5%	+10%	+20%	+50%
λ	0.383	0.366	0.368	0.371	0.363	0.367	0.364	0.371	0.371	0.367	0.375
Q	0.363	0.369	0.373	0.369	0.364	0.367	0.365	0.368	0.367	0.366	0.364
A	0.732	0.986	0.713	0.489	0.355	0.367	0.415	-	-	-	-
μ	0.367	0.361	0.361	0.366	0.368	0.367	0.369	0.369	0.373	0.373	0.393
H	0.369	0.368	0.371	0.377	0.376	0.367	0.369	0.364	0.369	0.370	0.367
f_1	0.368	0.366	0.367	0.366	0.371	0.367	0.371	0.368	0.369	0.371	0.368
P_1	0.369	0.368	0.370	0.369	0.368	0.367	0.368	0.370	0.368	0.368	0.365

Table 15: Out-of-sample MAE of the forecasts produced by the one-step IDNS having a particularly shocked set of initial parameters. The MAE calculation considers all horizons and all maturities. We highlight the values bellow the base MAE. *Note*: the deviations are applied to each of the matrices, vectors and standalone parameters one at a time: no shock is applied to two elements at the same time. In addition, there are missing values for the matrix A because a shock of +5% or above would breach the constraints on the autoregressive coefficients. Each value of the table is extracted after performing a backtest with the modified initial parameters as starting points of the MLE optimization.

Table 15 shows that the forecasts of the traditional one-step IDNS are somewhat sensible to the starting points of the optimization procedure. In general, for variations of between +/-1% and +/-50% in the value of the initial parameters, the forecast MAE does not vary more than a few hundredths of a unit. However, in the most extreme cases it varies up to several tenths of a unit, as observed in the case of negative variations of A. For most of the parameters, a whole set of different initial values can be used and the model will have a somewhat similar accuracy. An exception to this is the matrix A, whose starting values considerably impact the out-of-sample performance of the model. Indeed, a decrease of 20% in the value of A nearly triples the forecast error, while an increase of just 1% enlarges the forecast error by half a unit. These variations in the forecast error produced by a possible erroneous estimation of A are unmatched by the error variations obtained with the rest of the parameters. The out-of-sample forecast performance of the one-step IDNS is therefore considerably sensible to the starting value of A, more than any other parameter. As a consequence, special care should be put in estimating A correctly and making sure it contains reasonable values. Note that a secondary consequence of having further away initial parameter values is that it may induce a higher runtime.

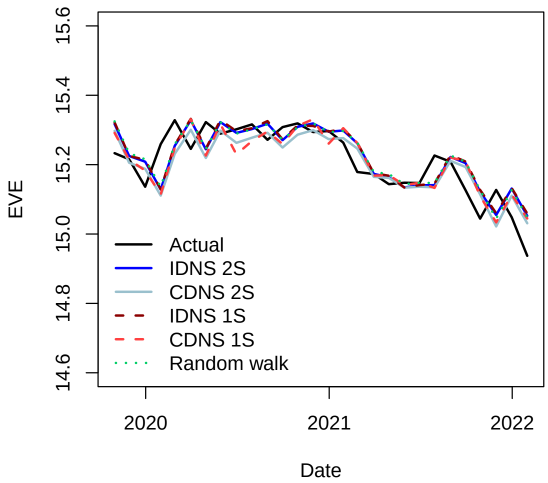
5 Impact of forecast discrepancies in the EVE of various simulated portfolios

In line with the motivation behind this work and to have a better idea of the magnitude of the discrepancies between the actual and forecasted interest rates, we calculate the EVE of three simulated portfolios using the interest rates obtained in our empirical assessment. What we refer to as portfolio is a set of future net cash-flows coming from interest rate sensible long or short positions. In this exercise, we assume all of our net positions are long, meaning that we assume that the cash-flows from the assets have been superior to the cash-flows from the liabilities. Our three portfolios are meant to represent three distinct scenarios: a scenario where the portfolio cash-flows are mostly concentrated on the short term (Portfolio ST), a scenario where the portfolio cash-flows are equally allocated between the short, medium and long term (Portfolio EQ), and finally a scenario where the portfolio cash-flows are concentrated on the long term (Portfolio LT). Given the characteristics of these three portfolios, models having superior forecasting capabilities for the long-end of the yield curve should be able to achieve better EVE forecasting performance for the LT portfolio, while a models having superior forecasting capabilities for the short-end of the yield curve should be able to achieve better EVE forecasting performance for the ST portfolio. Table 16 summarizes the cash-flows of each portfolio.

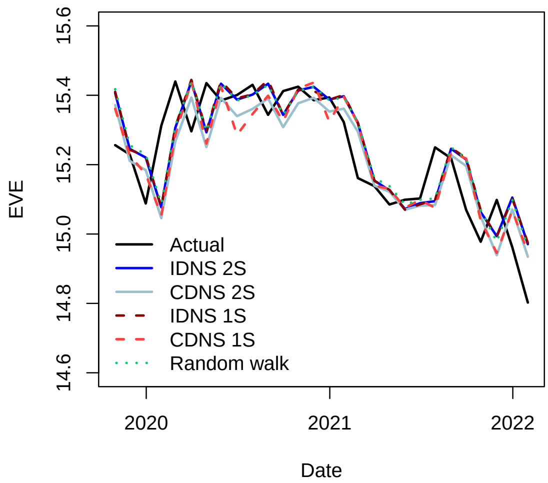
	Portfolio ST	Portfolio EQ	Portfolio LT
CF_{3M}	1.84	1.00	0.16
CF_{6M}	1.72	1.00	0.28
CF_{1Y}	1.60	1.00	0.40
CF_{2Y}	1.48	1.00	0.52
CF_{3Y}	1.36	1.00	0.64
CF_{4Y}	1.24	1.00	0.76
CF_{5Y}	1.12	1.00	0.88
CF_{6Y}	1.00	1.00	1.00
CF_{7Y}	0.88	1.00	1.12
CF_{8Y}	0.76	1.00	1.24
CF_{9Y}	0.64	1.00	1.36
CF_{10Y}	0.52	1.00	1.48
CF_{11Y}	0.40	1.00	1.60
CF_{12Y}	0.28	1.00	1.72
CF_{15Y}	0.16	1.00	1.84
Total	15	15	15

Table 16: Cash-flows of the ST, EQ and LT portfolios. *Note*: All cash-flows for each portfolio sum up to 15 monetary units.

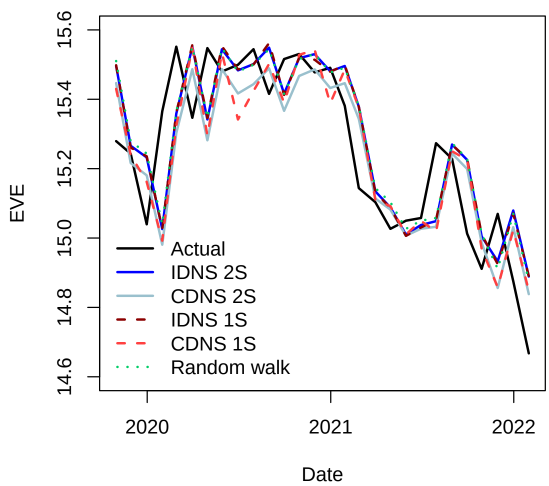
In the industry, it is also useful to forecast the EVE for a certain date in the future. For instance, a risk manager wanting to have an idea of what the value of the EVE is going to be a month, a quarter or a year from now. With this idea in mind, we forecast and compare the EVE of our three portfolios using the actual and forecasted interest rates from our empirical assessment for 1, 3, 6 and 12 months ahead. We assume the portfolio cash-flows are constant through time, meaning that we have the same portfolios for every forecast horizon. In that way we can isolate the effects of interest rate discrepancies. The results of this exercise are the following:



(a) ST portfolio

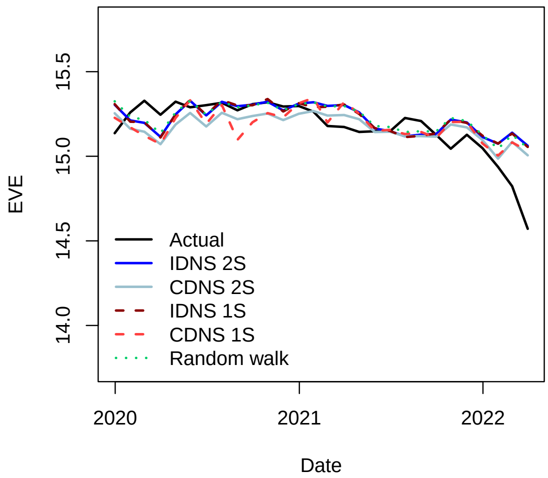


(b) EQ portfolio

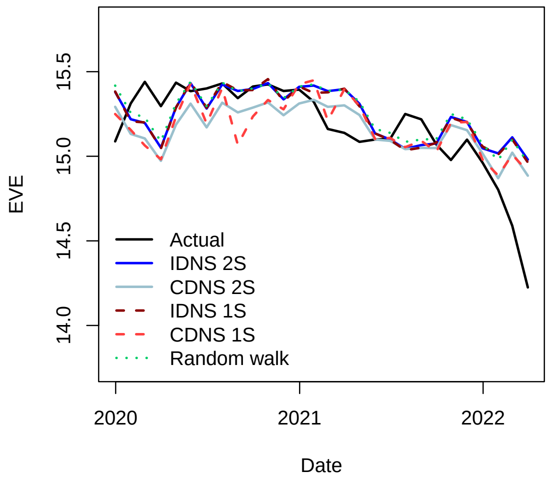


(c) LT portfolio

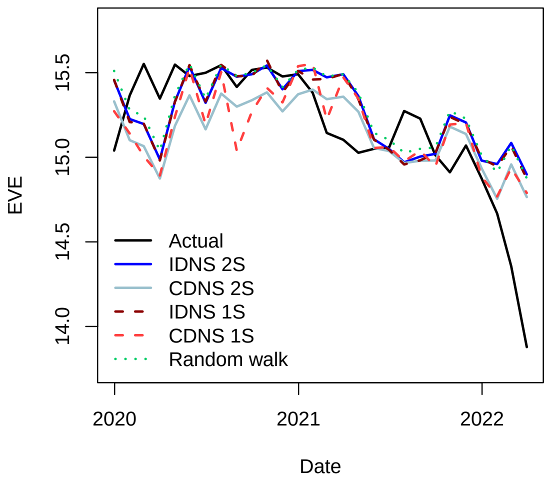
Figure 15: 1-month-ahead actual and forecasted EVE. *Note*: 2S stands for two-step and 1S stands for one-step.



(a) ST portfolio

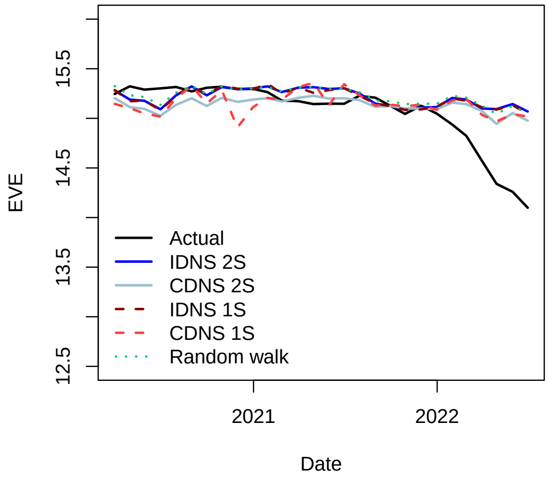


(b) EQ portfolio

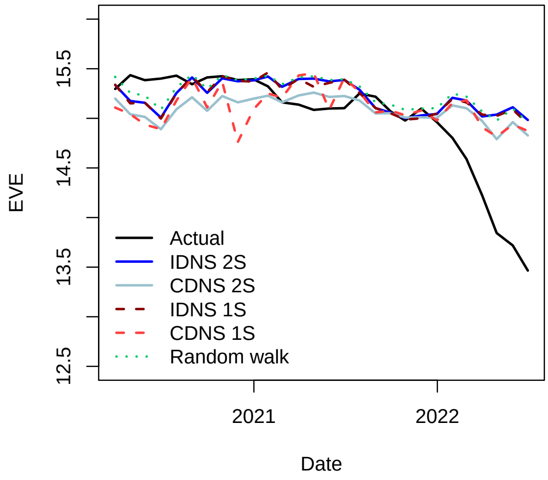


(c) LT portfolio

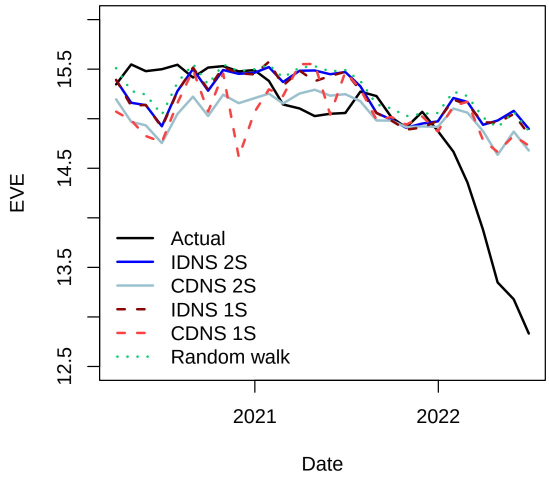
Figure 16: 3-month-ahead actual and forecasted EVE. *Note*: 2S stands for two-step and 1S stands for one-step.



(a) ST portfolio

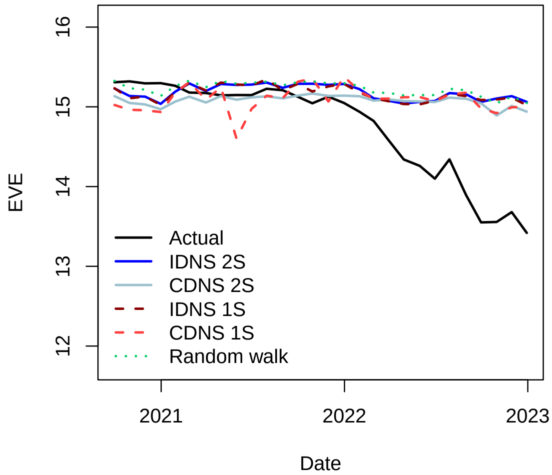


(b) EQ portfolio

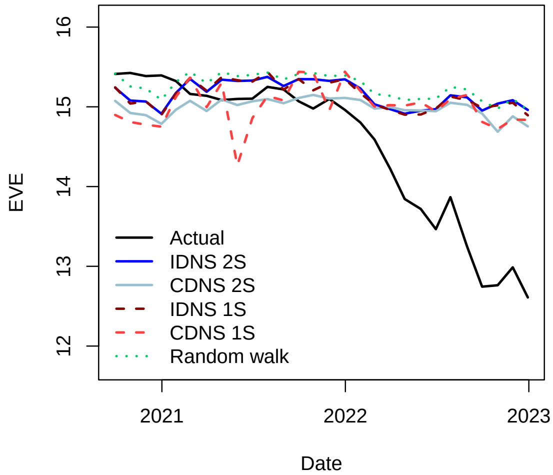


(c) LT portfolio

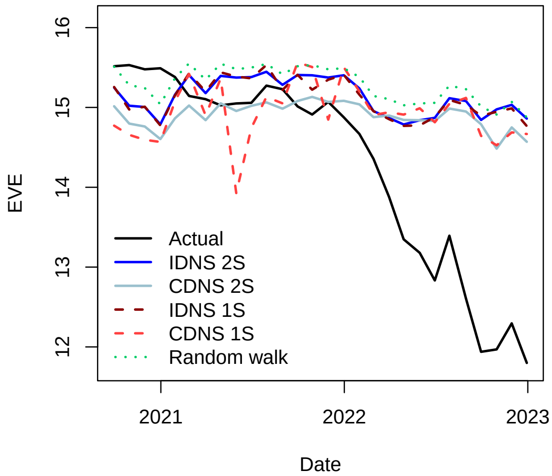
Figure 17: 6-month-ahead actual and forecasted EVE. *Note*: 2S stands for two-step and 1S stands for one-step.



(a) ST portfolio



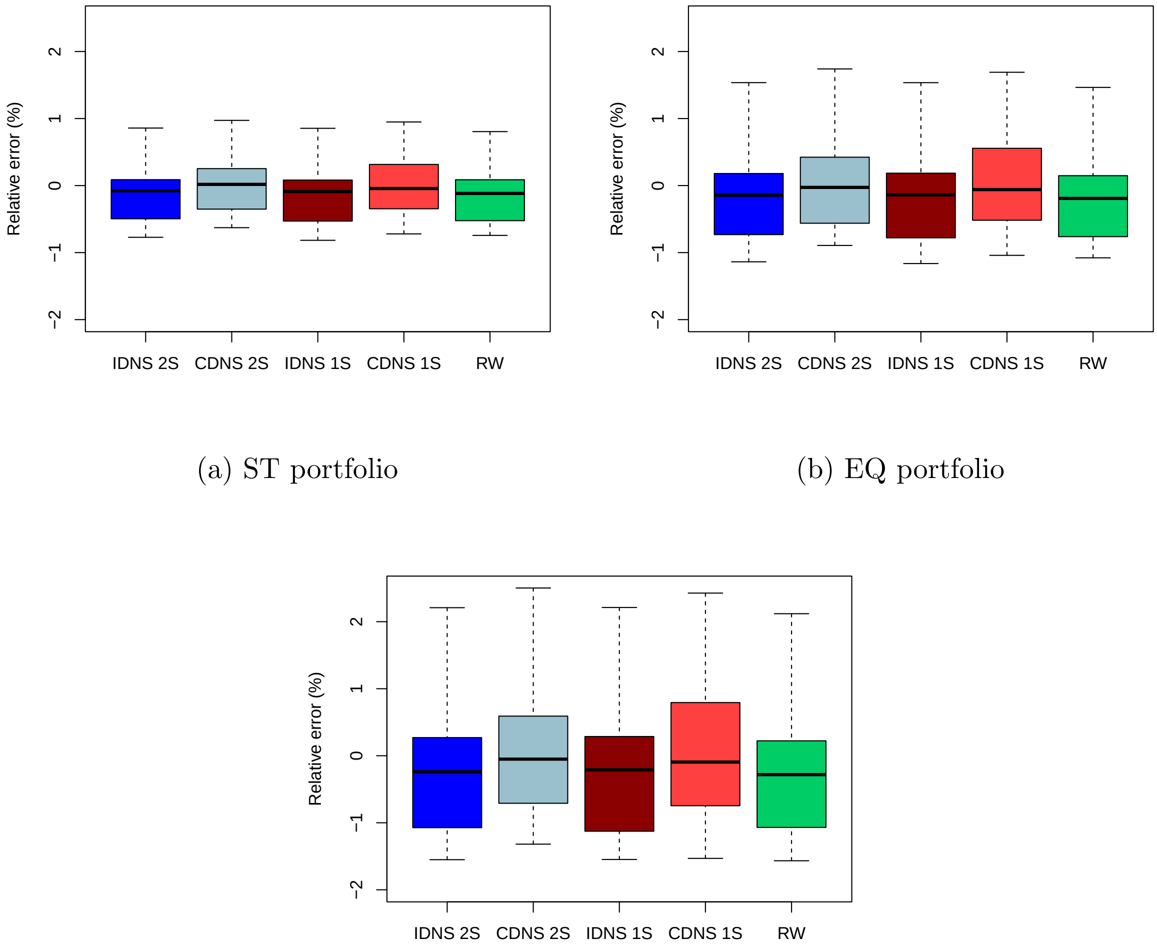
(b) EQ portfolio



(c) LT portfolio

Figure 18: 12-month-ahead actual and forecasted EVE. *Note*: 2S stands for two-step and 1S stands for one-step.

One can observe that at first the EVE (or equivalently the NPV) of the portfolios are above 15 monetary units even though the sum of their cash-flows are exactly equal to 15 monetary units. This is due to the negative interest rates present in the years 2020-2021. This was only temporary as interest rates increased by the end of 2021, producing a decrease in the portfolios' NPV, and therefore in their EVE. For short horizons, all forecasted EVE values sort of mimic the previous actual EVE because the autoregressive models have not yet had the time to diverge from a simple Random walk fashion estimation. Nonetheless, for longer horizons we see a more pronounced differentiation between the forecasts. Here are the relative errors between the actual and forecasted EVE values:

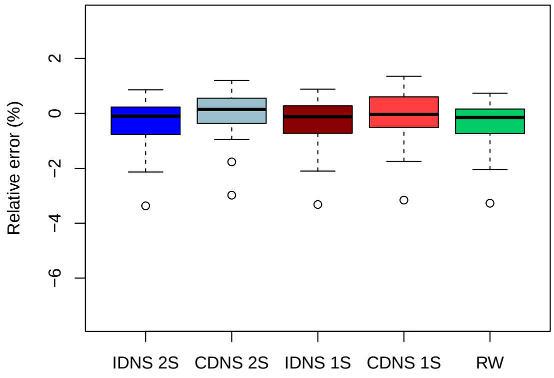


(a) ST portfolio

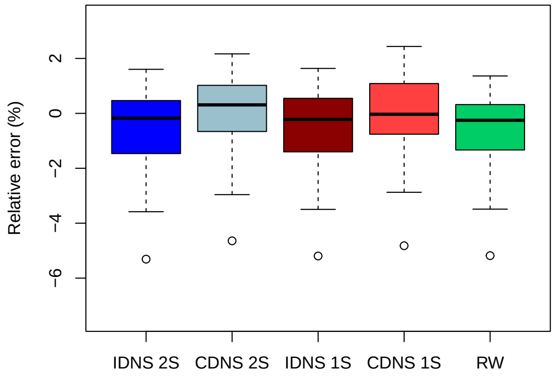
(b) EQ portfolio

(c) LT portfolio

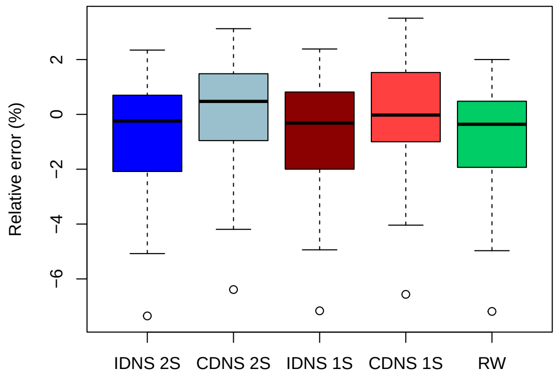
Figure 19: 1-month-ahead relative error (%) of the forecasted EVE through the backtest. *Note*: 2S stands for two-step, 1S stands for one-step, RW stands for Random walk.



(a) ST portfolio

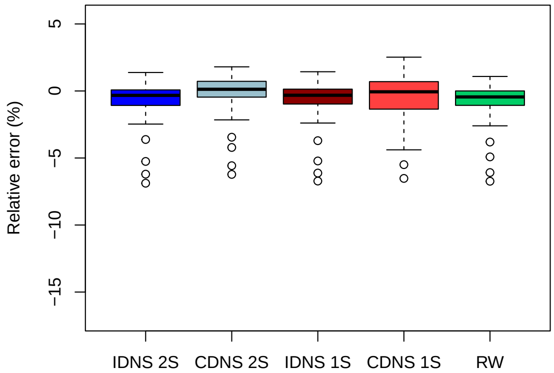


(b) EQ portfolio

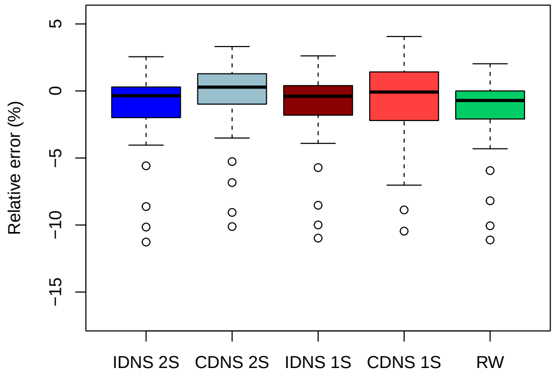


(c) LT portfolio

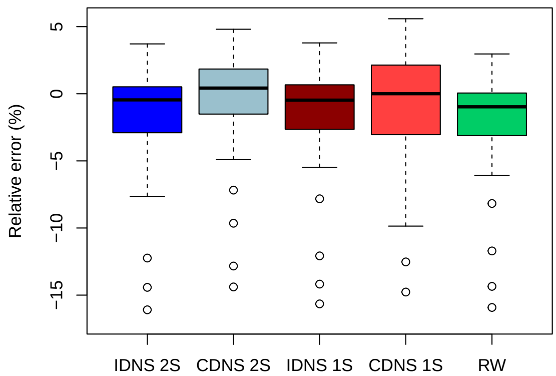
Figure 20: 3-month-ahead relative error (%) of the forecasted EVE through the backtest. *Note*: 2S stands for two-step, 1S stands for one-step, RW stands for Random walk.



(a) ST portfolio

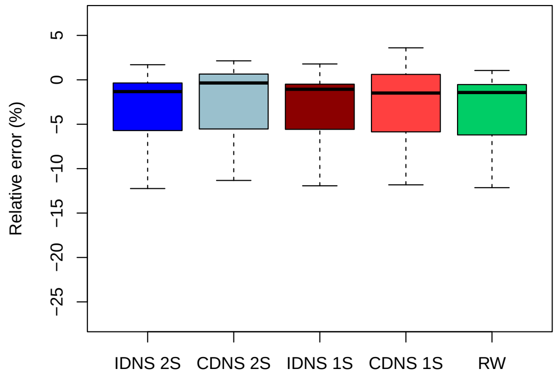


(b) EQ portfolio

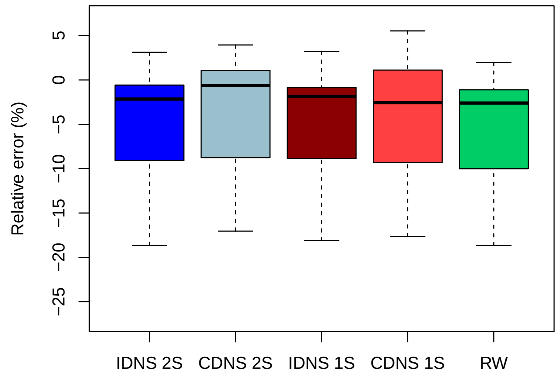


(c) LT portfolio

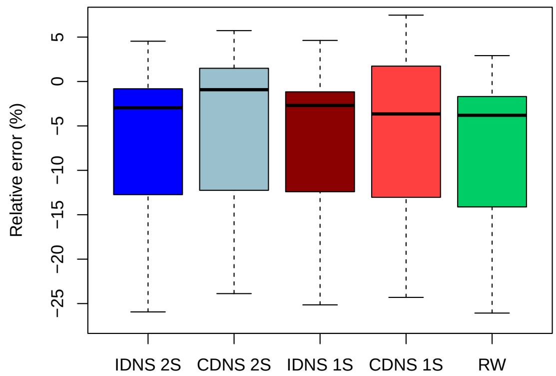
Figure 21: 6-month-ahead relative error (%) of the forecasted EVE through the backtest. *Note*: 2S stands for two-step, 1S stands for one-step, RW stands for Random walk.



(a) ST portfolio



(b) EQ portfolio



(c) LT portfolio

Figure 22: 12-month-ahead relative error (%) of the forecasted EVE through the backtest. *Note*: 2S stands for two-step, 1S stands for one-step, RW stands for Random walk.

For a 1 month horizon, the relative error ranges from around -2% to 2.5% in the most extreme case. However, as the horizon increases so does the error. Indeed, for 12 months ahead, the error ranges from around -25% to 5% in the most extreme case.

In this exercise, we also see that the apparent forecasting superiority of one model over another (as we could have concluded at first glance by the results in tables 11 and 12) becomes less apparent. For all of our simulated portfolios, the forecasted EVE of each model has essentially the same accuracy: their relative errors are quite similar inside each horizon as we can see on figures 19, 20, 21 and 22. In other terms, the superior forecasting accuracy of any given model in strict terms of yield curve forecasting becomes more marginal when it comes to the EVE calculation, even considering portfolios designed to accentuate differences in the forecasting capabilities on certain horizons and on certain parts of the yield curve. Of course, some differences still exist. For instance, the median of the errors can present differences between the models of around a tenth of a percent for 1-month-ahead forecasts up to around 3% in the most extreme case for the 12-months-ahead forecasts. We can observe that the distribution of the two-step CDNS relative errors are often a little more centered around 0 than the rest of the models. The decision of whether these differences are palpable enough will depend on the specific application of the EVE forecast, however it seems that the differences between models are not dramatic.

6 Conclusion and future work

Interest rate forecasting is an essential part of proper risk management of financial institutions. We have described the widespread Dynamic Nelson-Siegel model, evaluated its out-of-sample performance in the task of euro risk-free yield curve forecasting, analysed its robustness and evaluated the impact of the DNS forecast discrepancies on the EVE of three simulated portfolios. Regarding the forecasting performance, we found that the in-sample error minimization using the MAE as the error statistic is the criterion that allows for the highest out-of-sample accuracy for the two-step DNS. In addition, we show that the two-step CDNS generally outperforms the two-step IDNS, both the traditional one-step IDNS and CDNS and the Random Walk in several horizons and several maturities under our specific forecast framework. However, there is no clear winner between the two-step IDNS and the one-step IDNS. Considering both the IDNS and CDNS, there is no clear superiority of neither the two-step approach or the traditional one-step approach over the other. We also show that the Random walk does remain a competitive model and even outperforms the DNS model in many situations, notably for the 3-month-ahead forecasts. Finally, based on market data and simulated data, we demonstrate that the IDNS based on traditional one-step approach may produce more unstable parameter estimates than its counterparts. Not to mention that its implementation is more complex and prone to errors than the two-step approach. Despite the forecasting superiority of certain models in certain contexts, the differences between them don't seem to be enough to make a considerable difference when it comes to EVE forecasting performance.
Regarding the robustness assessment, we evaluated the sensibility of the two-step IDNS forecasting performance with respect to parameter λ. We found that λ can have a considerable impact on the out-of-sample performance of the two-step IDNS, specially for small values of λ. Lastly, we assessed the sensibility of the traditional one-step IDNS to its initial parameter values. We observe that the forecast errors stay within a relatively constant interval for variations of up to -/+ 50% of the value of the starting parameters. However, the forecasts remain somewhat unstable.

In terms of future work, an interesting comparative analysis could be done by expanding the traditional one-step procedure with Expectation-Maximization algorithms or even changing to a Bayesian approach. In addition, a useful analysis could be done by assessing the impact a variety of optimization algorithms can have on the stability of the parameter estimates via the one-step approach. Indeed, we found for instance that the change from the BFGS algorithm to the Nelder-Mead did produce substantial changes to the forecasting performance of our models.

More future work could be carried out by comparing the effectiveness of non-linear regression or even ridge regression to improve the two-step approach. Indeed, one could avoid fixing the shape parameter by using non-linear regression and reduce the instability of the estimated factors by using ridge regression.

Finally, extensions to the DNS model could be analysed in more detail. For instance, the Arbitrage-Free Nelson-Siegel is a popular alternative build on top of the DNS that takes care of the issued related to the arbitrage possibilities allowed by the DNS. The DNS with time-varying parameters has also been studied recently. As a last example, the integration of a GARCH model to model time-varying volatility could be analysed and compared.

7 References

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

Basel Committee on Banking Supervision. (2009). Range of practices and issues in economic capital frameworks.

Basel Committee on Banking Supervision. (2016). Interest rate risk in the banking book.

Basel Committee on Banking Supervision. (2019). Basel framework.

Bayazit, D. (2004). *Yield Curve Estimation and Prediction with Vasicek Model* (Master's thesis). Middle East Technical University.

Bernhart, G. (2013). *Interest rate bootstrapping explained* (tech. rep.). Xaia Investment GmbH.

Caldeira, J., & al. (2016). Forecasting the yield curve with the arbitrage-free dynamic Nelson-Siegel model: Brazilian evidence. *EconomiA*, *17*, 221–237.

Christensen, J.H.E., Diebold, X. and Rudebusch. G.D. (2011). *The Affine Argitrage-Free Class of Nelson-Siegel Term Structure Models* (Working Paper Series).

Clarke, J. (n.d.). *Constructing the OIS Curve* (tech. rep.).

De Pooter, M. (2007). *Examining the Nelson-Siegel Class of Term Structure Models* (Tinbergen Institute Discussion Paper).

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

Diebold, F., & Rudebusch, G. (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press.

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

European Banking Authority. (2022). Final report on guidelines on the management of interest rate risk and credit spread risk arising from non-trading book activities.

Filipović, D. (1999). A note on the Nelson-Siegel family. *Mathematical finance*, *Mathematical finance*.

Giaconomi, R., & White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, *74*(6), 1545–1578.

Huang, Z. (2021). *Fitting Yield Curve with Dynamic Nelson-Siegel Models: Evidence from Sweden* (Master's thesis). Uppsala Universitet.

Ibáñez, F. (2015). *Calibrating the Dynamic Nelson-Siegel Model: A Practitioner Approach* (tech. rep.). Central Bank of Chile.

Jungbacker, B., & Koopman, S. (2014). Likelihood-based Dynamic Factor analysis for Measurement and Forecasting. *Econometrics Journal*, *17*, 1–21.

Kim, Y., & Bang, H. (2018). Introduction to Kalman Filter and Its Applications. In *Introduction and Implementations of the Kalman Filter*. IntechOpen.

Koopman, S.J., Mallee, M.I.P. and van der Wel, M. (2007). Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters. *Journal of Business & Economic Statistics*, *2010*(3), 329–343.

Luethi, D., & al. (2022). *Package 'fkf'*.

Mamon, R. S. (2004). Three Ways to Solve for Bond Prices in the Vasicek Model. *Journal of Applied Mathematics and Decision Sciences*, *8*(1), 1–14.

Mishkin, F., & Serletis, A. (2015). *The economics of money, banking and financial markets, 4th edition*. Pearson.

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

Nymand-Andersen, P. (2018). *Yield Curve modelling and conceptual framework for estimating yield curves: evidence from the European Central Bank's yield curves* (Statistics Paper Series). European Central Bank.

Povala, P., & Vasil, R. (2017). *Forecasting The Term Structure of Interest Rates in Slovakia* (tech. rep.).

Shumway, R., & Stoffer, D. (2016). *Time Series Analysis and Its Applications, 4th edition*. Springer.

Toomet, O. (2021). *Maximum Likelihood Estimation with MaxLik* (tech. rep.).

Tsay, R. S. (2014). *Multivariate Time Series Analysis With R and Financial Applications*. Wiley.

8 Appendices

8.1 Forecasting procedure illustration

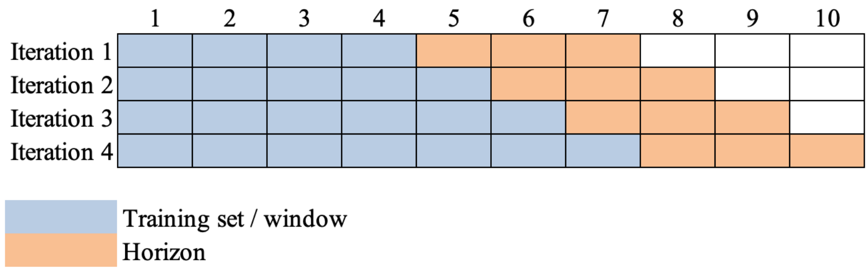


Figure 23: Illustration of an out-of-sample rolling origin forecast with constant holdout and expanding training set. In this illustrative case, the initial training set has 4 observations, the initial test set has 6 observations and the forecast horizon = 3.

8.2 Bootstrapping procedure

The objective of our bootstrapping procedure is to extract the implied yield curve from a market swap curve. In our case, we bootstrap the EUR OIS rates, using the ESTR overnight rate (we directly use ESTR for most recent dates and EONIA - 8.5 bps for dates when the ESTR was not published yet). We use the ACT/360 convention. Our procedure is based on the procedure from Bernhart, 2013.

Imagine a set of interest rate swaps (IRS) with tenors $T_0, T_1, ..., T_N$ where $T_i \in$ {overnight, 3M, 6M, 1Y, ..., 15Y}. Let s_i be the par swap rate of the IRS with tenor T_i. The swap rate s_0 is not quoted on the market but we keep it for a better understanding of the process. Let $r^{overnight}$ be the overnight rate underlying those IRS.

The bootstrapping procedure is the following:

1. Since the bootstrap is an iterative procedure going from small to big tenors, we calculate the overnight discount factor $D(0, T_0)$ (so the first discount factor in our series) using equation (35).
2. We use previously calculated discount factors in the swap rate equation to calculate each new discount factor one by one using equation (36).
3. Once we have all discount factors, we transform them into yields by using equation (37).

Overnight discount factor equation:

$$D(0, T_0) = \frac{1}{(1 + r^{overnight})^{\frac{1}{360}}} \tag{35}$$

Swap rate equation:

$$\begin{aligned} Fixed\ leg_i &= Floating\ leg_i \quad 1 \leq i \leq N \\ \Rightarrow \quad s_i \sum_{j=1}^{i} (T_j - T_{j-1}) D(0, T_j) &= 1 - D(0, T_i) \end{aligned}$$

$$\Rightarrow \quad D(0,T_i) = \frac{1 - s_i \sum_{j=1}^{i-1}(T_j - T_{j-1})D(0,T_j)}{1 + s_i(T_i - T_{i-1})} \tag{36}$$

Discount factor equation:

$$D(t,T) = e^{-(T-t)y(t,T)}$$

$$\Rightarrow \quad y(t,T) = \frac{-lnD(t,T)}{T-t} \tag{37}$$

8.3 Raw data inputs

EUR Overnight-Index-Swap rate data:

Source: Bloomberg

Frequency: Monthly

Price: Last traded price (PX Last)

From: 28/02/2007 (included)

To: 30/12/2022 (included)

Missing values: No

Maturity	Ticker
3M	EESWEC BGN Curncy
6M	EESWEF BGN Curncy
1Y	EESWE1 BGN Curncy
2Y	EESWE2 BGN Curncy
3Y	EESWE3 BGN Curncy
4Y	EESWE4 BGN Curncy
5Y	EESWE5 BGN Curncy
6Y	EESWE6 BGN Curncy
7Y	EESWE7 BGN Curncy
8Y	EESWE8 BGN Curncy
9Y	EESWE9 BGN Curncy
10Y	EESWE10 BGN Curncy
11Y	EESWE11 BGN Curncy
12Y	EESWE12 BGN Curncy
15Y	EESWE15 BGN Curncy

Table 17: Tickers from the EUR OIS.

ESTRON Index data:

Source: Bloomberg

Frequency: Monthly

From: 31/10/2019 (included)

To: 30/12/22 (included)

Missing values: No

EONIA Index data:

Source: Bloomberg

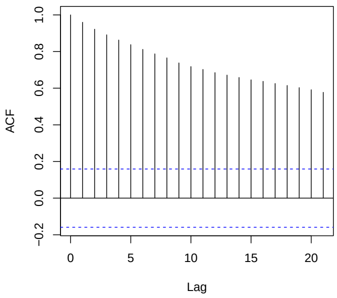
Frequency: Monthly

From: 28/02/2007 (included)

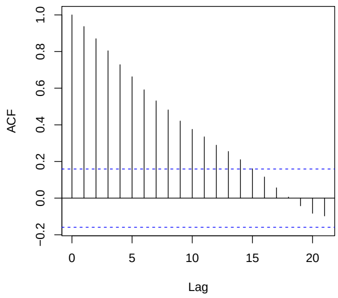
To: 31/12/21 (included)

Missing values: No

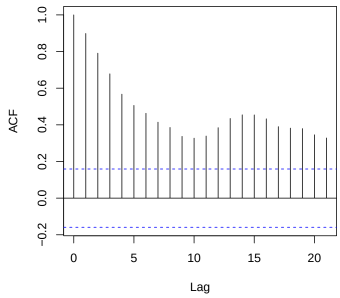
8.4 ACF and PACF of the factors derived from the initial training set yield curve via two-step approach



(a) Level factor ACF

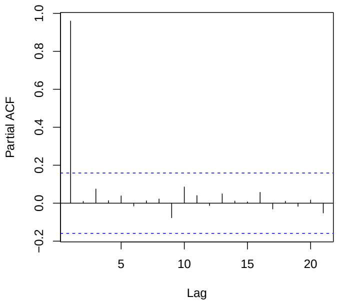


(b) Slope factor ACF

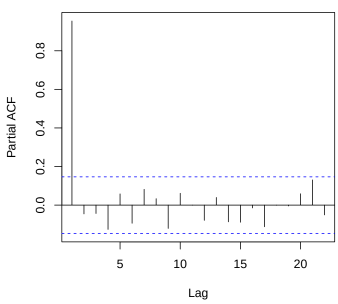


(c) Curvature factor ACF

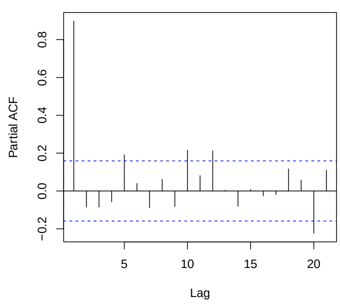
Figure 24: ACF of the initial training set yield curve factors.



(a) Level factor ACF



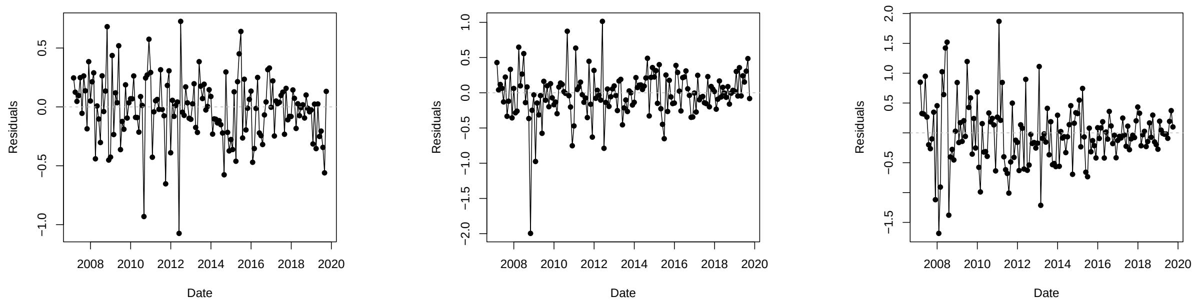
(b) Slope factor PACF



(c) Curvature factor PACF

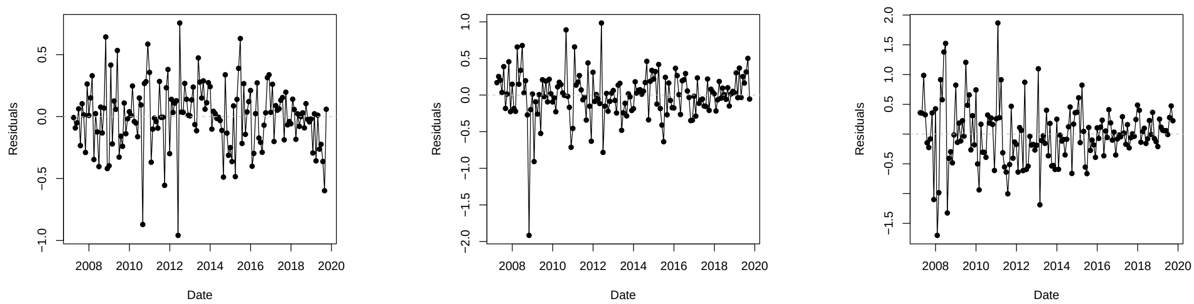
Figure 25: PACF of the initial training set yield curve factors.

8.5 In-sample residuals from the initial training set fit with the two-step DNS



(a) Level factor (b) Slope factor (c) Curvature factor

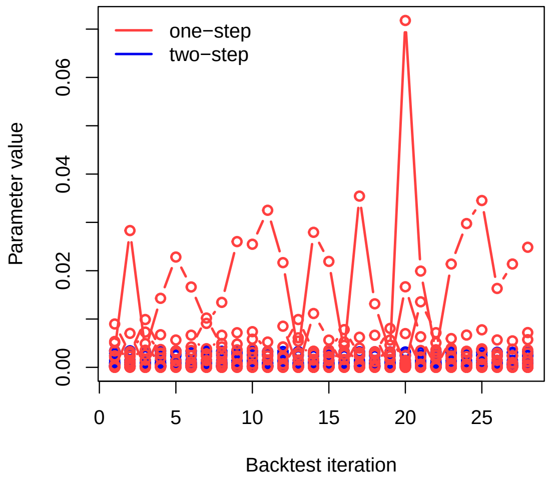
Figure 26: AR(1) in-sample residuals of the fitted factors for the initial training set yield curve.



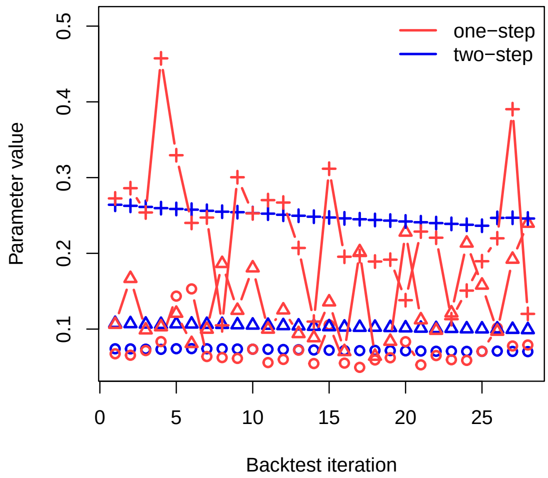
(a) Level factor (b) Slope factor (c) Curvature factor

Figure 27: VAR(1) in-sample residuals of the fitted factors for the initial training set yield curve.

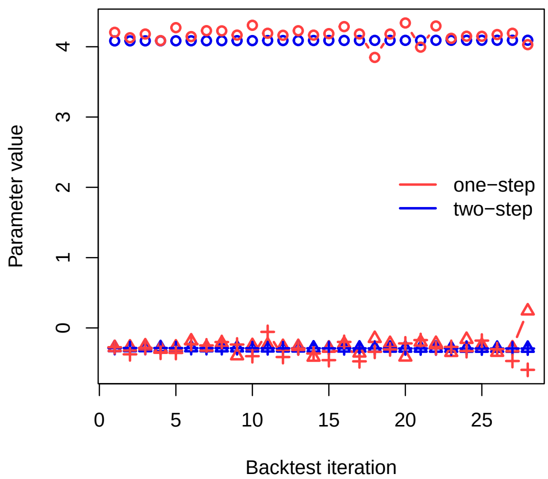
8.6 Estimated parameters at each iteration of the two-step vs one-step backtest



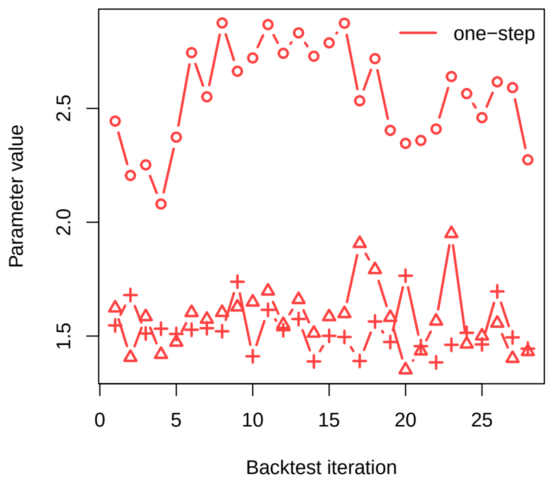
(a) H



(b) Circles: q_{11}, triangles: q_{22}, crosses: q_{33}



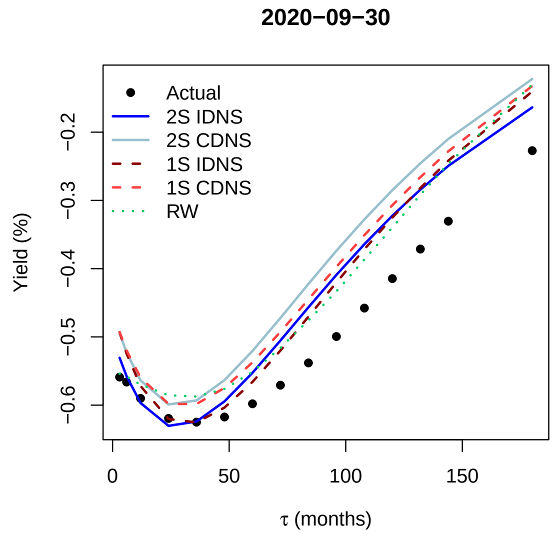
(c) Circles: l_1, triangles: s_1, crosses: c_1



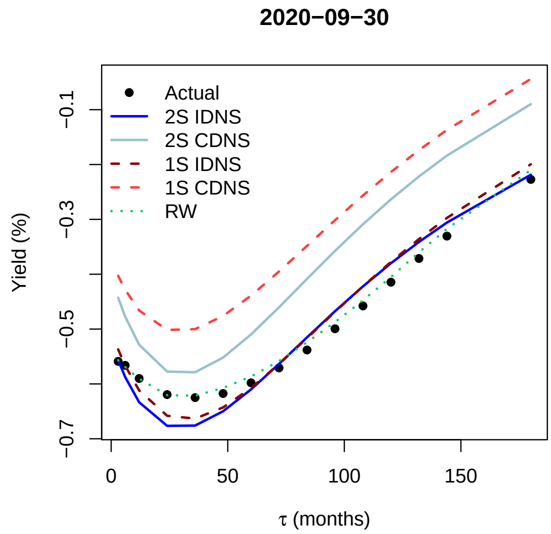
(d) Circles: P_{11}, triangles: P_{22}, crosses: P_{33}

Figure 28: Values of H, Q, f_1, P_1 for the two-step and one-step IDNS for each backtest iteration. The values of P_{ii} correspond to the diagonal elements of the initial error covariance matrix P_1.

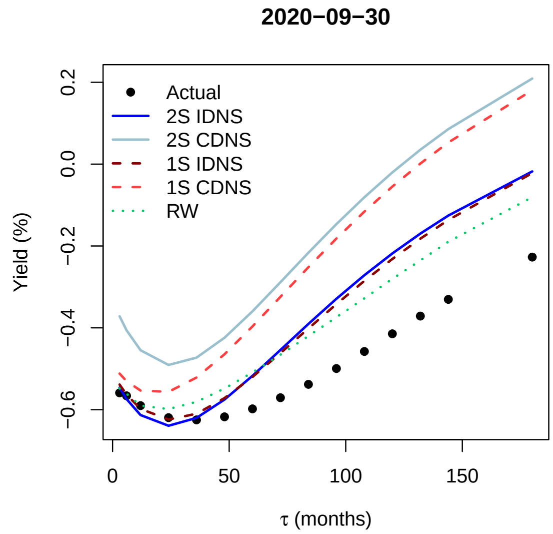
8.7 Out-of-sample fit of the competing models of the two-step vs one-step backtest



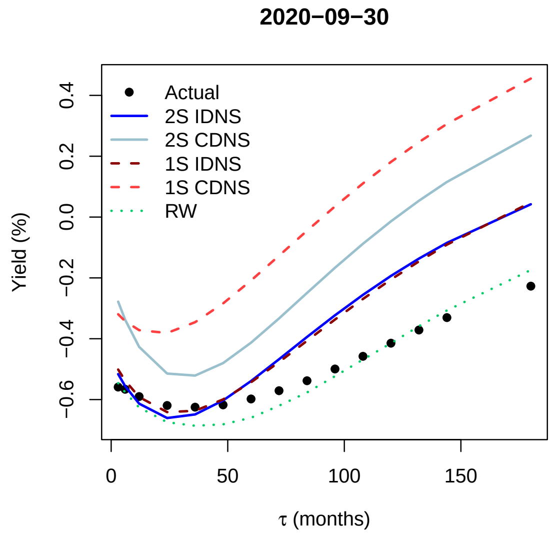
(a) $h = 1$ month ahead



(b) $h = 3$ months ahead

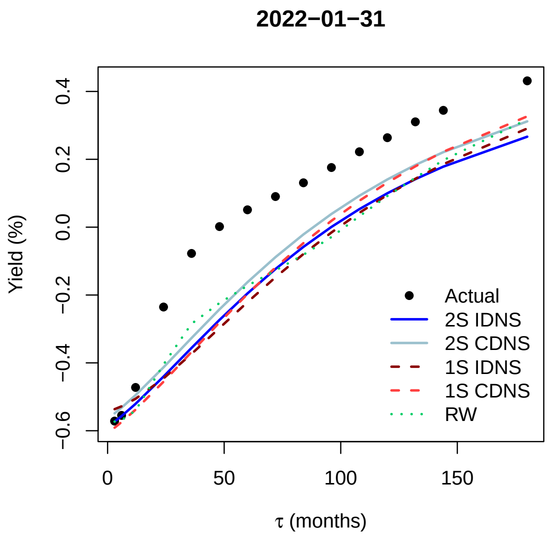


(c) $h = 6$ months ahead

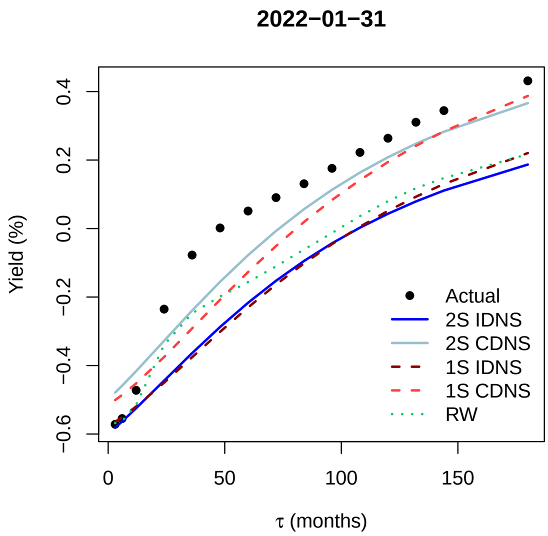


(d) $h = 12$ months ahead

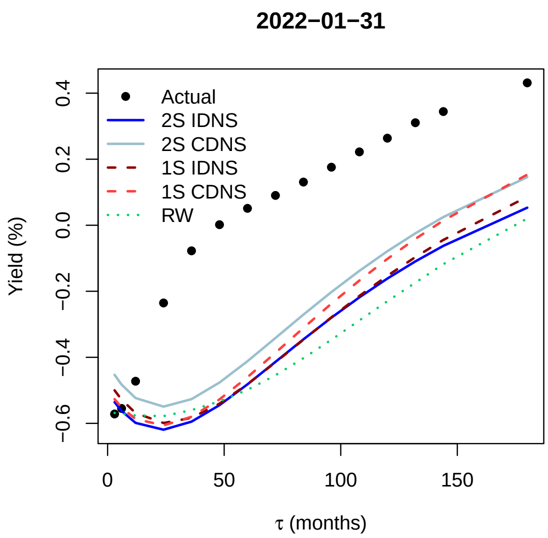
Figure 29: Out-of-sample forecasted yields and actual yields of the 30$^{\text{th}}$ of September 2020 for $h = 1,3,6,12$ months ahead. *Note*: 2S stands for two-step, 1S stands for one-step and RW stands for Random Walk.



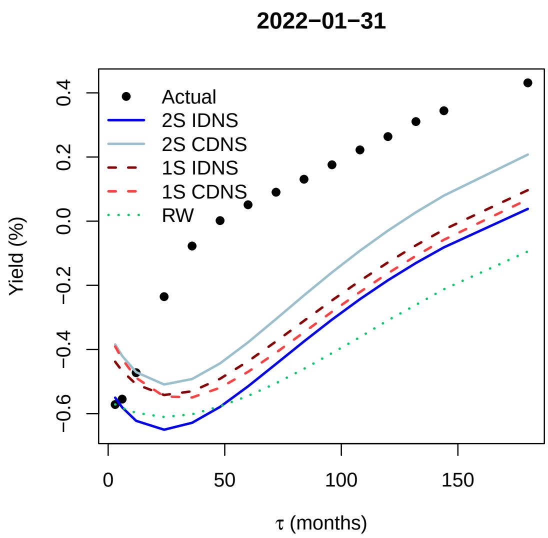
(a) $h = 1$ month ahead



(b) $h = 3$ months ahead



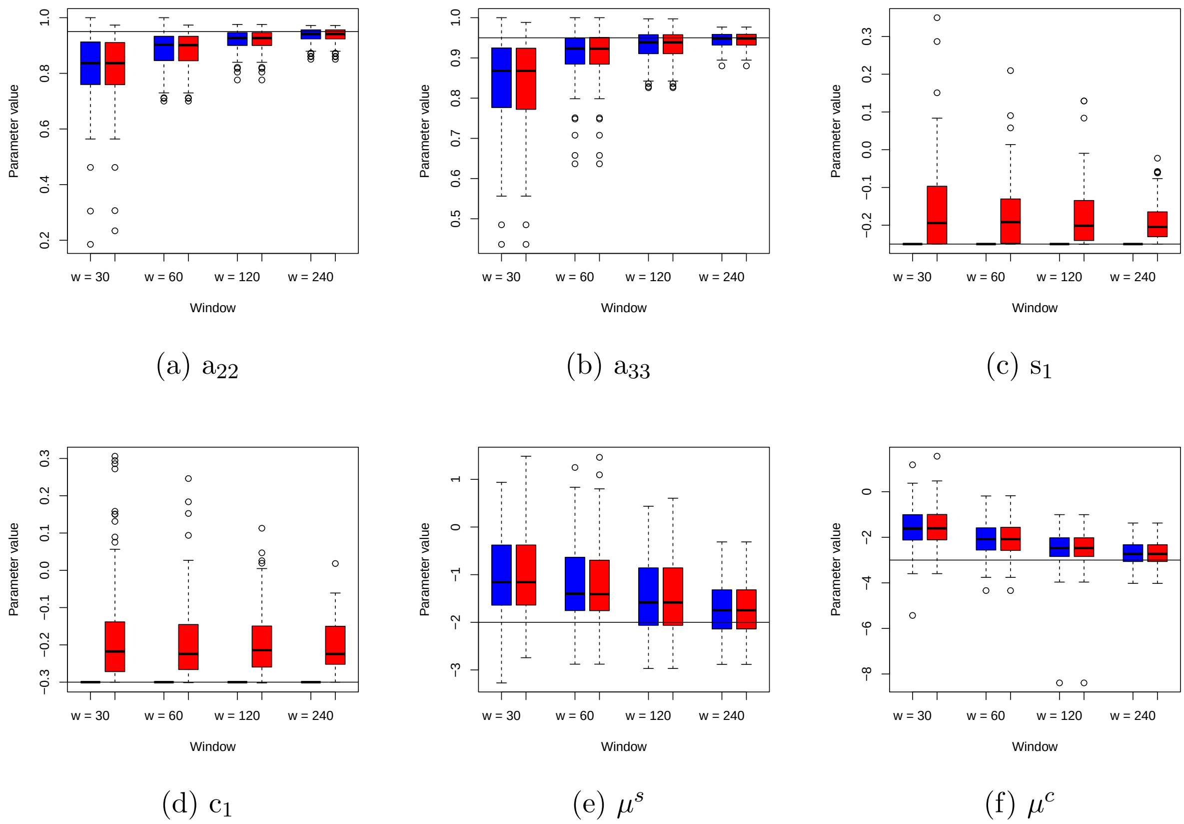
(c) $h = 6$ months ahead



(d) $h = 12$ months ahead

Figure 30: Out-of-sample forecasted yields and actual yields of the 1st of January 2022 for $h =$ 1,3,6,12 months ahead. *Note*: 2S stands for two-step, 1S stands for one-step and RW stands for Random Walk.

8.8 Estimated parameters in the simulation exercise



(a) a_{22} (b) a_{33} (c) s_1

(d) c_1 (e) μ^s (f) μ^c

Figure 31: Estimated parameters from 100 simulated yield curves for a given window. In blue: two-step IDNS, in red: one-step IDNS. The horizontal line is the actual value of the parameter.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Chaussée de Binche 151, 7000 Mons, Belgique | **www.uclouvain.be/lsm**