Modelling the Yield Curve in a Low Interest Rate Environment with a Markov-switching Dynamic Nelson-Siegel Model

M. Noteboom

414137

August 30, 2019

Erasmus School of Economics

Master Econometrics & Management Science

Specialisation: Quantitative Finance

Supervisor: prof.dr. M. van der Wel

Co-reader: dr. H.J.W.G. Kole

Abstract

This thesis introduces Markov-switching in the dynamic Nelson-Siegel model to fit and forecast the yield curve in a low interest rate environment. I allow for regime-switching in the mean of the Nelson-Siegel slope factor to distinguish between a period where the yield curve is flat and one where the curve is steep. I extend the regime-switching model by linking the yield curve to the macro-economy in three ways: by adding the macro-economic indicators as state variables; by letting the transition probabilities depend on the indicators; and a combination of both. The regime-switching model significantly improves the in-sample fit and out-of-sample forecasting performance relative to the regular dynamic Nelson-Siegel model. The addition of macro-economic indicators marginally increases the model fit. Out of all models, a regime-switching model that allows the transition probabilities to depend on the macro-economic indicators produces the most superior forecasts, especially at the short end of the yield curve.

Keywords: Nelson-Siegel yield curve model; State-space representation; Markov-switching; Zero lower bound; Kalman filter; Kim filter; Macro-finance

*The views stated in this thesis are those of the author and do not necessarily reflect those of the Erasmus School of Economics or Erasmus University. I thank my supervisor prof.dr. Van der Wel and co-reader dr. Kole for helpful comments. All remaining errors are mine.

Contents

1 Introduction

The term structure of interest rates, also known as the yield curve, is the relation between interest rates or bond yields and different terms or maturities. Interest rates play several pivotal roles within the economy. One of them is being a key policy instrument of central banks to steer the state of the economic environment. It is therefore of great interest of various market participants such as investors, policy makers, and risk managers to accurately predict interest rates movements. For investors, forecasting interest rates may result in higher portfolio returns. However, for policy makers understanding the change in future interest rates may help their decision making concerning macro-economic monetary policy.

Many of the currently used dynamic term structure models (DTSMs) are not able to properly capture the behavior of the yield curve in periods of low interest rates. This is because DTSMs assume that the probability of yields falling close to or below zero is negligible (see e.g. Bauer and Rudebusch (2016)). However, during the global financial crisis in 2008, the US nominal short rate has dropped toward 0% - 0.25% and has remained at historically low levels ever since. When interest rates are almost at their zero lower bound (ZLB), the dynamics of the fixed-income market change due to three factors. First, the yield curve exhibits asymmetric yield movement since a downside shift of the yield curve is far less likely than an upward shift (see e.g. Christensen (2013)). Second, monetary institutions are reluctant to decrease interest rates when the yield curve is near the ZLB.[1] Third, investors always have the option to extract their money from the bank and hold cash which might be more profitable during periods of extreme low interest rates. Therefore, it is necessary to consider alternative ways of modelling the term structure in a low interest rate environment.

The growing literature on modelling the yield curve during periods of low interest rates can be split into two branches. A first approach is based on the Gaussian shadow-rate model of Black (1995) which sets the short-term interest rate equal to the maximum of zero (that is, the ZLB) and a shadow interest rate. The shadow interest rate is an alternative Federal funds rate that, contrary to the nominal short rate, can become negative.[2] The shadow rate reflects the policy of the Federal Reserve (Fed) during unconventional monetary policies.[3] Gorovoi and Linetsky (2004) show that analytical solutions exist for a one-factor shadow-rate model. However, due

[1] In order not to lower interest rates near the ZLB, central banks have to rely on unconventional forms of monetary policy. Examples include purchasing long-term assets (better known as quantitative easing) or publicly communicating about the current state of the economy and possible future monetary policy (better known as forward guidance).

[2] In fact, the nominal short rate has attained negative values in the past. This is due to the non-negligible costs of transacting and holding large amounts of cash. For instance, in mid-2016 the 1-year German bund paid a negative interest rate of -0.60%.

[3] See Black (1995) and Wu and Xia (2016) for an elaborate discussion on shadow interest rates.

to the computational burden of estimating multi-factor shadow-rate models, the literature has only considered one- or two-factor models.

The second branch of literature is much more nascent and is based on modelling the term structure with a regime-switching model. Hevia et al. (2015) characterize the regimes by a latent Markov-switching component and use the dynamic Nelson-Siegel (DNS) model of Diebold and Li (2006) to incorporate a regime-switching decay parameter. They motivate their modelling approach by arguing that the shape of the yield curve heavily depends on the stance of monetary policy. Christensen (2013) introduces regime-switching in the arbitrage-free Nelson-Siegel model of Christensen et al. (2011) and distinguishes between two states: a state with normal interest rate levels, and one where interest rates are extremely low. Both studies find that the proposed model forecasts better than a single-regime model and that the regime-switching is related to business cycles.

This thesis follows the second strand of literature and uses a Markov-switching dynamic Nelson-Siegel (MS-DNS) model to capture the behaviour of the yield curve in a low interest rate environment. I use the model of Bernadell et al. (2005) and distinguish between a regime with normal interest rate levels, and one where interest rates are low. In the MS-DNS model, regime-switching occurs in the mean of the Nelson-Siegel slope factor. This is because during periods of economic downturn, central banks often lower short-term interest rates to stimulate the economy. As long-term interest rates most likely remain stable or rise, lowering short-term rates causes an increase in the slope of the yield curve. However, when inflation levels are high, central banks try to moderate the economy by for example increasing the short-term rate. Hence, this thesis evaluates whether the MS-DNS model performs better than the single-regime DNS model, in terms of model fit and forecasting performance.

I extend the MS-DNS model by linking the shape of the yield curve to the macro-economy. A direct way to study the effect of macro-economic indicators on the term structure is by including them in the set of state variables, as in Diebold et al. (2006) (DRA). However, a more subtle approach is to let the indicators enter the MS-DNS model through the transition probabilities. In this set-up, the state of the indicators determine the transition probabilities in each period, resulting in time-varying transition probabilities. Lastly, I combine both of the approaches and let the macro-economic indicators enter the model through the state variables as well as through the time-varying transition probabilities. It is worthwhile to consider a joint macro-finance perspective because the dynamics of the yield curve and the state of the macro-economy are jointly related. For instance, the short-term interest rate is based on the policy rate of central banks which is a key instrument used for monetary policy.[4] According to the Taylor

[4]Since long-term interest rates can be viewed as a weighted average of expected future short-term rates, one

(1993) rule, the decisions of central banks also heavily rely on macro-economic indicators such as the current inflation level and the gross domestic product.[5] Therefore, I use the policy rate, the current inflation level, and the gross domestic product to study to what extent the addition of macro-economic indicators improves the performance of the MS-DNS model.

Since interest rates were far away from zero before the crisis in 2008, Bernadell et al. (2005) mainly used the MS-DNS model to distinguish between periods where the shape of the yield curve differs (e.g. during different stances of monetary policy). However, the MS-DNS model is also suitable for modelling the ZLB period. For instance, after the financial crisis it was not possible for the American central bank to implement an expansionary monetary policy since interest rates were already extremely low. Therefore, the Fed adopted its quantitative easing program to fight domestic deflation by buying government bonds and riskier assets with a longer maturity. This caused a rapid increase in liquidity in the market and a rise in the supply of money in the economy. The liquidity premium charged on these securities decreased which was reflected in the form of reduced medium term interest rates and, hence, a steep yield curve.[6]

In the regime-switching models, regimes are characterized by a latent Markov-switching component. This component determines which state drives the dynamics of the three latent Nelson-Siegel factors, better known as the level, slope and curvature, respectively. Kim and Nelson (1999) introduce an efficient way to estimate the latent factors while simultaneously determine the current state, called the Kim filter. In essence, the Kim filter combines both the Kalman filter to extract the latent factors and the Hamilton filter to compute the state probabilities. The in-sample fit of the regime-switching models relative to the baseline DNS model is studied by considering their residual diagnostics. However, the true residuals in the regime-switching models are unobserved due to the latent Markov-switching component. Hence, I adopt the Rosenblatt (1952) transformation to ensure that the residuals of the DNS and regime-switching models are compared on a fair basis. I also present goodness-of-fit statistics and perform likelihood ratio tests to study the in-sample performance of the models. The forecasts constructed by the regime-switching models are compared with those of the DNS model and a random walk.

Based on US Treasury yield data, the MS-DNS model is indeed able to distinguish between a regime with normal and low interest rate levels. The mean of the slope factor is slightly negative in the first regime indicating a nearly flat yield curve. In the second regime, the mean

may even argue that the whole yield curve responds to changes in the macro-economy.

[5]Several studies, including DRA, indeed include these indicators to examine the performance of the macro DNS model. In the context of regime-switching models, Zhu and Rahman (2009) use the state-space representation of DRA to extend the DNS model to a regime-switching macro-finance model.

[6]For an extensive study on the effects of the quantitative easing program on interest rates, I refer to Krishnamurthy and Vissing-Jorgensen (2011).

is, however, considerably more negative which corresponds to a steep curve. In terms of model fit, I find that the MS-DNS model is superior relative to the baseline DNS model. The regime-switching model is particularly able to model the short end of the yield curve better, except for the 3-month yield. The filtered state probabilities coincide quite well with NBER recessions and periods where the short-term interest rate is relatively low, such as the ZLB period. The MS-DNS model also generates better forecasts than the DNS model and a random walk at the short end of the curve.

The inclusion of macro-economic indicators only marginally improves the in-sample performance of the MS-DNS model. Out of the three considered approaches to include the indicators, allowing the transition probabilities to depend on the indicators and, therefore, vary over time is the most promising. In terms of forecasting performance, all macro regime-switching models forecast the short end of the yield curve better than the MS-DNS model. The macro model with time-varying transition probabilities, again, outperforms the other two macro models. Including macro-economic indicators through both the state variables and the transition probabilities does not further improve the in-sample fit and out-of-sample forecasting performance relative to either one of the two approaches.

This thesis contributes towards the growing literature on regime-switching term structure modelling by incorporating macro-economic indicators in the model of Bernadell et al. (2005). As opposed to Bernadell et al. (2005) who a priori determine the regimes by subjective threshold levels based on economic indicators, I endogenously link the shape of the yield curve to the macro-economy. I also introduce a macro regime-switching model with macro-economic state variables as well as macro-dependent transition probabilities, and compare its performance against a model which only considers one of the two approaches to include the indicators. Lastly, this thesis extends the paper of Bernadell et al. (2005) by evaluating the out-of-sample performance of the (macro) regime-switching models during the recent periods of extreme low interest rates in the US.

The remainder of the thesis is organized as follows. Section 2 describes the baseline DNS model and its regime-switching extensions. Section 3 describes the data on US Treasury yields and macro-economic indicators. In Section 4 and 5, I present and discuss the in-sample fit and forecasting performance of the Nelson-Siegel models, respectively. Concluding remarks and directions for future research are given in Section 6.

2 Models and Methodology

This section introduces the Nelson-Siegel framework. In Section 2.1, I present the model of Nelson and Siegel (1987). Section 2.1.1 discusses the baseline DNS model and Section 2.1.2 presents its state-space representation. Section 2.1.3 elaborates on the estimation of the DNS model which is based on the Kalman filter. In Section 2.2, I introduce regime-switching in the DNS model and extend this model by including macro-economic indicators. Section 2.3 discusses the likelihood ratio test which is used to statistically test for the presence of the second regime.

2.1 The Nelson-Siegel Model

The original Nelson and Siegel (1987) model gives a static description of the yield curve in the form of a factor model. In the corresponding representation, the term structure is expressed in terms of a small set of unobserved factors:

$$y(\tau_i) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} \right) + \beta_3 \left(\frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} - e^{-\lambda \tau_i} \right), \tag{2.1}$$

where $y(\tau_i)$ is the yield of a bond maturing in τ_i months, β_1, β_2 and β_3 are latent factors, and λ is a loading parameter, for $i = 1, ..., N$ different maturities. Due to the simplicity of the Nelson-Siegel yield curve model, many central banks have adopted Nelson-Siegel type of models to fit the term structure of interest rates, as described in the documentation from the Bank for International Settlements (1999).

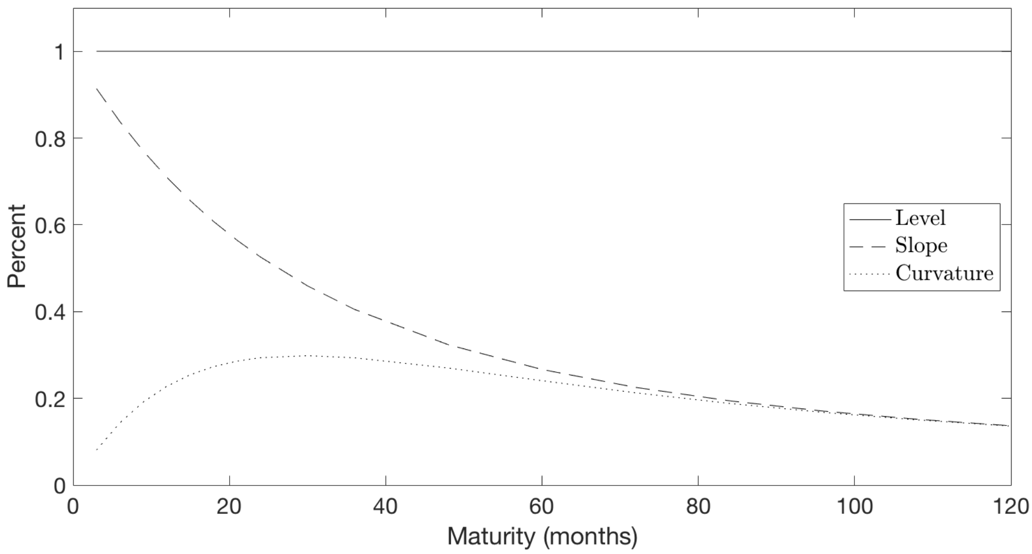
2.1.1 The Dynamic Nelson-Siegel Model

A major drawback of the Nelson-Siegel model, as specified in (2.1), is its poor out-of-sample forecasting performance. This is because the latent Nelson-Siegel factors tend to vary over time. Hence, Diebold and Li (2006) dynamically extend the Nelson-Siegel model by allowing for time-varying latent factors at each time t. This gives the DNS model:

$$y_t(\tau_i) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda_t \tau_i}}{\lambda_t \tau_i} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_t \tau_i}}{\lambda_t \tau_i} - e^{-\lambda_t \tau_i} \right). \tag{2.2}$$

In the DNS model, β_{1t}, β_{2t} and β_{3t} determine the dynamics of $y_t(\tau_i)$, while the expression next to each factor, the factor loading, describes the cross-section of the yields. Diebold and Li (2006) show that β_{1t}, β_{2t} and β_{3t} can be interpreted as a level, slope and curvature factor, respectively. This interpretation becomes more clear in Figure 1, which shows a plot of the factor loadings as a function of the maturity. For instance, the loading on β_{1t} is constant and equal to one for all maturities. An increase in β_{1t}, hence, results in an equal increase in yields across all maturities. Therefore, the loading on β_{1t} may be interpreted as a long-term or level factor.

Figure 1: Nelson-Siegel factor loadings



Note: This figure shows the loadings of the Nelson-Siegel level, slope and curvature factor as a function of maturity for a decay parameter $\lambda = 0.0609$; the value found by Diebold and Li (2006).

To interpret the loading on β_{2t} and β_{3t}, it is useful to consider the limits of the DNS model in (2.2). These are given by

$$\lim_{\tau_i \to 0} y_t(\tau_i) = \beta_{1t} + \beta_{2t}, \qquad \lim_{\tau_i \to \infty} y_t(\tau_i) = \beta_{1t}. \tag{2.3}$$

Since the loading on β_{2t} only affects the short end of the yield curve, it is often interpreted as a short-term factor. This factor is also closely related to the slope of the yield curve, which can be defined as the yield on an infinitely long bond minus the yield on an infinitely short bond and is equal to $-\beta_{2t}$. Hence, Diebold and Li (2006) interpret the short-term factor as a slope factor.[7] Finally, from (2.3) it is evident that the loading on β_{3t} does not affect the short end nor the long end of the term structure. However, an increase in β_{3t} does affect medium-term yields since the corresponding factor loading, $\left(\frac{1-e^{-\lambda_t \tau_i}}{\lambda_t \tau_i} - e^{-\lambda_t \tau_i}\right)$, is an increasing function for middle-term maturities. Therefore, this factor is also called the medium-term or curvature factor.

In the DNS model, the time-varying loading parameter λ_t determines the exponential decay rate. For instance, small values of λ_t correspond to a slow decay of the yield curve and result in a better fit of the curve at long maturities. Conversely, large values of λ_t fit the curve better at short maturities. Moreover, λ_t determines the maturity where the loading on the curvature factor β_{3t} is maximized.

Since the DNS formulation in (2.2) is a nonlinear function of the parameters, the set of parameters $\boldsymbol{\theta} = \{\beta_{1t}, \beta_{2t}, \beta_{3t}, \lambda_t\}$ can be estimated with nonlinear least squares, for each month

[7] Following the literature however, I define β_{2t} as the slope factor instead of $-\beta_{2t}$ for ease of discussion.

t. However, a potential disadvantage of this estimation procedure is that one has to rely on iterative optimization procedures to compute $\boldsymbol{\theta}$ which heavily depend on the chosen starting values for the unknown parameters. Hence, it is of interest to consider alternative ways to deal with the decay parameter λ_t. Koopman et al. (2010) propose to model λ_t as a fourth time-varying unobserved component, and use the extended Kalman filter to estimate the DNS model. However, I follow Diebold and Li (2006) and fix λ_t in the DNS model.[8] This greatly facilitates the estimation procedure as the DNS formulation in (2.2) is now linear in the unknown parameters β_{1t}, β_{2t} and β_{3t}.

2.1.2 The DNS Model as a State-Space Representation

In addition to extending the Nelson-Siegel framework by allowing for time-varying latent factors, Diebold and Li (2006) also find that the factors are highly correlated over time. Therefore, they suggest to model the dynamics of the factors with either a multivariate VAR(1) process or three univariate AR(1) processes. This allows one to easily estimate the DNS model with the following two-step procedure. First, estimate β_{1t}, β_{2t} and β_{3t} in equation (2.2) for each yield curve at time t by cross-sectional OLS. Afterwards, one can use the (V)AR(1) process to estimate the factors at time $t+1$. Introducing an autoregressive process in the DNS model also aids in forecasting future yields. The obtained estimates of the latent factors in the second step of the two-step approach can simply be plugged into equation (2.2) to generate bond yields at time $t+1$.

DRA recognize that the DNS model with an autoregressive process directly forms a state-space model. They introduce the following state-space representation for a VAR(1) specification:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & & \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix}, \tag{2.4}$$

where

$$\begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{pmatrix}. \tag{2.5}$$

Here, equation (2.4) is referred to as the measurement equation, which describes the cross-section of yields with the three latent factors β_{1t}, β_{2t}, and β_{3t}. Equation (2.5) is also called the transition

[8] From now on, I drop the subscript and refer to the decay parameter as λ.

equation and models the dynamics of the factors. The corresponding state-space system can easily be re-written into vector notation as

$$\boldsymbol{y}_t = \boldsymbol{X}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}), \tag{2.6}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta}_{\boldsymbol{t-1}} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}), \tag{2.7}$$

where $\boldsymbol{y}_t = (y_t(\tau_1), ..., y_t(\tau_N))'$ is the $(N \times 1)$ vector of yields, $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$ the (3×1) vector of latent factors, $\boldsymbol{X}$ the $(N \times 3)$ matrix of factor loadings and $\boldsymbol{F}$ the (3×3) matrix of autoregressive coefficients.[9] An advantage of the DNS state-space representation is that all parameters can be estimated simultaneously with the Kalman filter. For instance, De Pooter (2007) finds that the use of the Kalman filter gives better out-of-sample results than the aforementioned two-step approach. This is because the latter procedure does not take into account the estimation error in the transition equation when estimating the measurement equation.

For optimality of the filter, it is necessary that the measurement error $\boldsymbol{\epsilon}_t$ and transition disturbance $\boldsymbol{\eta}_t$ are white noise. Furthermore, both $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are required to be orthogonal to each other and to the unobserved factors $\boldsymbol{\beta}_t$. I also assume that the covariance matrix $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_\eta$ are diagonal. Following DRA, the assumption of a diagonal $\boldsymbol{\Sigma}_\epsilon$ matrix is quite common since it indicates that deviations of implied yields from the yield curve are not correlated between different maturities. Opting for a diagonal $\boldsymbol{\Sigma}_\eta$ matrix allows the shocks to the factors in $\boldsymbol{\beta}_t$ to be uncorrelated over time.

In previous literature, the AR(1) specification of the DNS model has been preferred over the VAR(1) specification for several reasons. For instance, Christensen et al. (2011) show that the off-diagonal elements in the autoregressive matrix $\boldsymbol{F}$ are not statistically significant. Diebold and Li (2006) also argue that the VAR(1) forecasts are inferior to the AR(1) forecasts since unrestricted VARs tend to produce poor forecasts of economic variables and are subject to overfitting. For completeness, I estimate both versions of the DNS model and choose one of the two model specifications based on the estimation results.

2.1.3 Estimation of the DNS Model

Since the DNS model assumes a fixed decay parameter λ, the measurement equation in (2.6) becomes linear in the latent factors $\boldsymbol{\beta}_t$. Consequently, the DNS model in (2.6) and (2.7) becomes a linear Gaussian state space model. Following DRA, I therefore use the Kalman filter in combination with Maximum Likelihood (ML) to estimate $\boldsymbol{\beta}_t$. The Kalman filter is a recursive procedure that computes estimates of $\boldsymbol{\beta}_t$ based on I_t, the available information at time t. In

[9] One obtains the DNS state-space representation for three univariate AR(1) processes by assuming the matrix $\boldsymbol{F}$ to be diagonal.

the first step, one chooses starting values for $\boldsymbol{\beta}_t$ and its covariance matrix $\boldsymbol{P}_t$. These are then used in the prediction and updating step of the Kalman filter to form optimal predictions of $\boldsymbol{\beta}_t$ and $\boldsymbol{P}_t$ given I_t. The other parameters in the DNS model are collected in the parameter set $\boldsymbol{\theta}$ and are assumed to be unknown. Optimal parameter values $\boldsymbol{\theta}_{ML}$ are obtained by numerically maximizing the log-likelihood function. Below, I describe each step of the Kalman filter and ML in more detail.

Initialisation

Recall that the state-space representation of the DNS model in vector notation is given by

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{X}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim N(0, \boldsymbol{\Sigma}_\epsilon), \\ \boldsymbol{\beta}_t &= \boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(0, \boldsymbol{\Sigma}_\eta). \end{aligned}$$

Given this model set-up, the set of parameters to be estimated is $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{F}, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_\eta, \lambda\}$. As starting values for $\boldsymbol{\mu}, \boldsymbol{F}$, and $\boldsymbol{\Sigma}_\eta$, I use the estimates of the DNS model obtained from the two-step approach described in Section 2.1.2. Furthermore, I initialise λ with the value found by Diebold and Li (2006), which is 0.0609, and set $\boldsymbol{\Sigma}_\epsilon$ equal to the identity matrix such that the yield deviation is equal for every maturity.[10] Lastly, the state vector $\boldsymbol{\beta}_{0|0}$ is initialised with its unconditional mean, $\boldsymbol{\beta}_{0|0} = (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{\mu}$, and its covariance matrix $\boldsymbol{P}_{0|0}$ with its unconditional covariance matrix, $\text{vec}(\boldsymbol{P}_{0|0}) = (\boldsymbol{I} - \boldsymbol{F} \otimes \boldsymbol{F})^{-1}\,\text{vec}(\boldsymbol{\Sigma}_\eta)$.[11]

Prediction step

In the prediction step, one uses the initial parameter values $\boldsymbol{\theta}^{(0)}$ as input in the Kalman filter to obtain an estimate of $\boldsymbol{\beta}_{t|t-1}$, the latent factors at time t given I_{t-1}, and its covariance matrix $\boldsymbol{P}_{t|t-1}$. The prediction step consists of the following two equations:

$$\boldsymbol{\beta}_{t|t-1} = \boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta}_{t-1|t-1}, \tag{2.8}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{F}\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{\Sigma}_\eta. \tag{2.9}$$

[10]Recall that λ maximizes the loading on the medium-term factor, which is usually taken to be a zero-coupon bond with a two- or three-year maturity. By taking the average of both maturities, which is 30 months, Diebold and Li (2006) find that the value of λ that maximizes the loading on the medium-term factor equals 0.0609.
[11]For the derivation of the unconditional mean and covariance matrix, I refer to Appendix A.

Updating step

When $\boldsymbol{y}_t$ is realized at the end of time t, one can calculate the prediction error $\boldsymbol{\nu}_{t|t-1}$. This prediction error contains new information about $\boldsymbol{\beta}_t$ that is not contained in $\boldsymbol{\beta}_{t|t-1}$. Therefore, one can make a more accurate inference on $\boldsymbol{\beta}_t$ after observing $\boldsymbol{y}_t$, which is based on all information up to time t. The updating step of the Kalman filter is summarized in the following four steps:

$$\begin{aligned}
\boldsymbol{\nu}_{t|t-1} &= \boldsymbol{y}_t - \boldsymbol{y}_{t|t-1} = \boldsymbol{y}_t - \boldsymbol{X}\boldsymbol{\beta}_{t|t-1}, \\
\boldsymbol{f}_{t|t-1} &= \boldsymbol{X}\boldsymbol{P}_{t|t-1}\boldsymbol{X}' + \boldsymbol{\Sigma}_\epsilon, \\
\boldsymbol{\beta}_{t|t} &= \boldsymbol{\beta}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{X}'\boldsymbol{f}_{t|t-1}^{-1}\boldsymbol{\nu}_{t|t-1}, \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{X}'\boldsymbol{f}_{t|t-1}^{-1}\boldsymbol{X}\boldsymbol{P}_{t|t-1},
\end{aligned}$$

where $\boldsymbol{f}_{t|t-1} = E[\boldsymbol{\nu}_{t|t-1}^2]$, the conditional variance of the prediction error.

Maximum Likelihood

After the prediction and updating step, the output of the Kalman filter is used to obtain an estimate of the parameter set $\boldsymbol{\theta}^{(1)}$. This is done via ML by maximizing the log-likelihood function. If one assumes that the conditional distribution of $\boldsymbol{y}_t$ is normal, one can use the prediction-error decomposition to obtain the relevant predictive log-likelihood as

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\ln(2\pi|\boldsymbol{f}_{t|t-1}|) - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\nu}_{t|t-1}'\boldsymbol{f}_{t|t-1}^{-1}\boldsymbol{\nu}_{t|t-1}, \tag{2.10}$$

which is maximized with respect to $\boldsymbol{\theta}$. The new estimates of the unknown parameters $\boldsymbol{\theta}^{(1)}$ serve as input for the prediction and updating step in the next iteration of the Kalman filter. One obtains optimal parameter values $\boldsymbol{\theta}_{ML}$ when these parameters maximize the log-likelihood function in equation (2.10).

2.2 The Markov-switching Dynamic Nelson-Siegel Model

In this section, I extend the baseline DNS model by allowing for regime-switching. Section 2.2.1 presents the MS-DNS model of Bernadell et al. (2005) which lets the DNS model switch between two regimes. The first regime is a period with normal interest rate levels and a nearly flat yield curve. In the second regime, short-term rates are considerably lower indicated by a steep curve. Afterwards, I propose several approaches to include macro-economic indicators in the MS-DNS model in Section 2.2.2. Section 2.2.3 discusses the estimation of the regime-switching models which is based on the Kim filter.

2.2.1 A Yield Curve Model without Macro-Economic Indicators

Most of the empirical literature on term structure modelling does not assume distinct regimes within the term structure. However, the existence of regimes is intuitively appealing since regimes can be linked to different states of the economy and, hence, have immediate impact on the shape of the yield curve. Studies such as Ang and Bekaert (2002) and Dai et al. (2007) also find considerable support for the existence of regime-switching in interest rates.

I adjust the DNS model by introducing regime-switching in the mean of the slope factor β_{2t}. The unobserved Markov component S_t, the realized state at time t, follows a two-state Markov process with transition probability matrix P given by

$$P = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix},$$

where $p_{ij} = P[S_t = j | S_{t-1} = i]$. Furthermore, S_t is equal to 1 when the yield curve is nearly flat (e.g. the short end is away from the ZLB) or 2 when it is steep (e.g. the short end is near the ZLB). In the resulting MS-DNS model of Bernadell et al. (2005), the measurement equation is the same as in (2.6), but the transition equation is modified to

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(0, \boldsymbol{\Sigma}_\eta), \tag{2.11}$$

where

$$\boldsymbol{\mu}_{S_t} = \begin{pmatrix} \mu_1 \\ \mu_{2,S_t} \\ \mu_3 \end{pmatrix}, \qquad \boldsymbol{F} = \begin{pmatrix} f_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & f_{33} \end{pmatrix}.$$

Similar to the DNS model, I assume that the white noise measurement error $\boldsymbol{\epsilon}_t$ and transition disturbance $\boldsymbol{\eta}_t$ are uncorrelated to each other and to the factors, and that $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_\eta$ are diagonal. It is also possible to include the slope factor, f_{22}, in the autoregressive matrix $\boldsymbol{F}$. However, Bernadell et al. (2005) state that under this specification regime-switching hardly occurs and that the states are far less persistent. Hence, I follow Bernadell et al. (2005) and set f_{22} equal to zero.

The motivation for modelling the regime-behaviour in this way is based on the Taylor rule and the Fisher decomposition. Taylor (1993) discusses how central banks, for example the Fed, should adjust interest rates as shifts in the economy occur, such as periods with lower than expected GDP growth or higher inflation levels. For example, when economic growth is falling (in terms of GDP), the Fed should try to stimulate the economy by lowering short-term interest rates. As long-term rates are unaffected by the monetary policy of the Fed, lowering short-term rates causes an increase in the slope of the yield curve. On the other hand, during periods of high inflation, the Fed should moderate the economy by raising its policy rate, the Federal funds rate, which is a short-term rate. This increases all other short-term rates and causes the yield curve to flatten.

The assumption that the long-term rate is more stable than the short-term rate rests on the principle of the Fisher decomposition. According to the Fisher decomposition, the nominal interest rate consists of the sum of the expected real interest rate and the inflation rate. It is known in macro-economics that, in the long run, the expected real rate is equal to the growth of the economy. However, in the short run the expected real rate is also affected by other factors causing the real rate to be more volatile. Therefore, the long-term rate is expected to be more stable and is, hence, unaffected by monetary policy decisions.[12]

2.2.2 A Yield Curve Model with Macro-Economic Indicators

I extend the MS-DNS model by linking the shape of the yield curve to the macro-economy. A straightforward approach is proposed by DRA who include k macro-economic indicators to the set of state variables $\boldsymbol{\beta}_t$. However, since in the MS-DNS model the off-diagonal elements of the transition matrix $\boldsymbol{F}$ are set to zero, there is no linkage between the macro-economic indicators and latent factors. Therefore, I modify $\boldsymbol{F}$ as follows

$$\boldsymbol{F}^* = \left(\begin{array}{c|c} \text{Diagonal}_{3\times3} & \text{Unrestricted}_{3\times3} \\ \hline 0_{k\times k} & \text{Diagonal}_{k\times k} \end{array}\right), \tag{2.12}$$

where the upper left block consists of the Nelson-Siegel factors (and the slope factor is still set to zero), and the bottom right block includes the k indicators. In this way, the yields do not give any feedback to the macro-economic indicators but the indicators do affect the yields through the Nelson-Siegel latent factors.[13] The corresponding MS-DRA model is similar to the MS-DNS model except that the dimensions are increased and $\boldsymbol{F}$ is replaced by $\boldsymbol{F}^*$:

$$\boldsymbol{y}_t = \boldsymbol{X}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}), \tag{2.13}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{F}^*\boldsymbol{\beta}_{\boldsymbol{t-1}} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}), \tag{2.14}$$

[12]The implication of the Fisher decomposition that long-term rates are more stable than short-term rates is also empirically found in yield curve data and is, hence, considered to be one of the stylized facts of the yield curve.

[13]The modification of the transition matrix $\boldsymbol{F}$ in this manner is first introduced by Exterkate et al. (2013) who extend the DNS model by including factors extracted from a large set of macro-economic indicators.

where $\boldsymbol{\beta}_t$, $\boldsymbol{\mu}_{S_t}$ and $\boldsymbol{\eta}_t$ are $(3+k) \times 1$, and $\boldsymbol{F}^*$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ are $(3+k) \times (3+k)$. Furthermore, $\boldsymbol{X}$ is $(3+k) \times (3+k)$ where the k rightmost columns contain only zeros such that the yields load only on the Nelson-Siegel factors. This specification of $\boldsymbol{X}$ is consistent with the general idea that the level, slope and curvature are sufficient to capture almost all of the cross-sectional variance of the yields (see Litterman and Scheinkman (1991)).

A downside of the two previous regime-switching models is that the transition probabilities are assumed to be constant over time; that is, the probability of switching from one regime to the other does not depend on the behaviour of underlying economic fundamentals. A constant transition probability may lead to model misspecification for estimating the yield curve. For example, given a current regime with low interest rate levels, the probability the economy stays within this regime decreases over time. Changes in leading economic indicators may also cause central banks to adjust their policy rates which in turn affect the shape of the yield curve. Therefore, macro-economic indicators may contain valuable information, particularly for the evolution of the transition probabilities over time.

I consider time variability in the transition probabilities by introducing a logit model, as proposed by Diebold et al. (1994). This transformation ensures that the transition probabilities are bounded between zero and one. The time-varying transition probabilities (TVTPs) are modeled as

$$p_{jj,t} = P[S_t = j | S_{t-1} = j, \boldsymbol{z}_{t-1}] = \frac{\exp(\boldsymbol{z}'_{t-1}\boldsymbol{\gamma}_j)}{1 + \exp(\boldsymbol{z}'_{t-1}\boldsymbol{\gamma}_j)}, \qquad j \in \{1, 2\}, \tag{2.15}$$

where the TVTP matrix of S_t is given by

$$P_t = \begin{pmatrix} p_{11,t} & 1 - p_{11,t} \\ 1 - p_{22,t} & p_{22,t} \end{pmatrix}. \tag{2.16}$$

Here, $\boldsymbol{z}_{t-1} = (1, z_{1,t-1}, ..., z_{k,t-1})'$ and is a $(k+1) \times 1$ vector containing macro-economic indicators that affect the state transition probabilities. Furthermore, $\boldsymbol{\gamma}_j = (\gamma_j^0, \gamma_j^1, ..., \gamma_j^k)'$ and is a $(k+1) \times 1$ vector of parameters governing the transition probabilities in regime j and which determine the weights of the macro-economic indicators in $\boldsymbol{z}_{t-1}$. The first element, γ_j^0, serves as a constant and provides additional fit in modelling the TVTPs. The inclusion of a constant also reduces the model with TVTPs to the MS-DNS model (with fixed transition probabilities) when the indicators are not significant, that is, $\gamma_j^1 = \gamma_j^2 = ... = \gamma_j^k = 0$. The MS-DNS model in (2.6) and (2.7) together with the TVTP model in (2.15) forms the MS-TVTP model.

A last extension of the MS-DNS model combines both of the previous approaches: the macro-economic indicators enter the model through the transition matrix $\boldsymbol{F}^*$ and through the transition probabilities in (2.15). In this way, the indicators affect the yields within each regime through the state variables, as well as across each regime through the TVTPs. I refer to the MS-DRA model in (2.13) and (2.14) together with the TVTP model in (2.15) as the MS-DRA-TVTP model.

2.2.3 Estimation of the MS-DNS Model

To estimate the MS-DNS model, one needs to use the Kalman filter to extract estimates of the latent factors, and the Hamilton filter to determine the state and transition probabilities. Kim and Nelson (1999) develop such a procedure where the Hamilton filter is embedded within the Kalman filter, better known as the Kim filter. The last step of the Kim filter, the collapsing step, ensures that the dimension of the Kalman filter stays tractable. At each iteration of the Kim filter, the parameter exhibiting regime-switching, in the MS-DNS model $\boldsymbol{\mu}_{S_t}$, is updated through a weighting scheme where the weights are estimated by the Hamilton filter. Optimal parameter values $\boldsymbol{\theta}_{ML}$ are obtained from repeating the Kim filter and ML until the log-likelihood function is maximized. Below, I explain each step of the Kim filter and ML in more detail, which closely follows the Kim filtering as described in Kim and Nelson (1999).

Initialisation

In the MS-DNS model, the set of parameters to be estimated is $\boldsymbol{\theta} = \{\boldsymbol{\mu}_{S_t}, \boldsymbol{F}, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_\eta, \lambda, p_{11}, p_{22}\}$. The initialisation of the Kim filter is similar to that of the Kalman filter described in Section 2.1.3; that is, I use the two-step estimates of the DNS model as starting values for $\boldsymbol{\mu}_1 = (\mu_1, \mu_{2,1}, \mu_3)'$, $\boldsymbol{F}$, and $\boldsymbol{\Sigma}_\eta$. To ensure different estimates of $\mu_{2,1}$ and $\mu_{2,2}$, I initialise $\mu_{2,2}$ as two plus the two-step estimate of $\mu_{2,1}$. Furthermore, the decay parameter λ is set to 0.0609 and $\boldsymbol{\Sigma}_\epsilon$ to the identity matrix. The transition probabilities p_{11} and p_{22} are set to 0.98 and 0.90, respectively, to ensure persistence within each regime. I initialise the state vector in regime j, $\boldsymbol{\beta}^j_{0|0}$, with its unconditional mean, $\boldsymbol{\beta}^j_{0|0} = (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{\mu}_{S_t}$, and its covariance matrix $\boldsymbol{P}_{0|0}$ with its unconditional covariance matrix, $\text{vec}(\boldsymbol{P}_{0|0}) = (\boldsymbol{I}-\boldsymbol{F}\otimes\boldsymbol{F})^{-1}\,\text{vec}(\boldsymbol{\Sigma}_\eta)$. Here, it should be noted that the covariance matrix $\boldsymbol{P}_{0|0}$ does not depend on regime j since the regime-switching enters the MS-DNS model through the mean of the transition equation $\boldsymbol{\mu}_{S_t}$. Finally, I use the steady-state probabilities as initial values for the state probabilities $P[S_0 = j|I_t]$, that is

$$P[S_0 = 1] = \frac{1 - p_{11}}{2 - p_{11} - p_{22}},$$
$$P[S_0 = 2] = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}.$$

Kalman filter

In the MS-DNS model, one is not interested in an estimate of $\boldsymbol{\beta}_t$ given I_t but also conditional on the state S_t being in regime j at time t and in regime i at time $t-1$

$$\boldsymbol{\beta}_{t|t-1}^{(i,j)} = E[\boldsymbol{\beta}_t | I_{t-1}, S_t = j, S_{t-1} = i],$$

where its covariance matrix $\boldsymbol{P}_{t|t-1}$ is defined similarly. Conditional on $S_{t-1} = i$ and $S_t = j$, the Kalman filter described in Section 2.1.3 is modified as follows

$$\boldsymbol{\beta}_{t|t-1}^{(i,j)} = \boldsymbol{\mu}_j + \boldsymbol{F}\boldsymbol{\beta}_{t-1|t-1}^{i}, \tag{2.17}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{F}\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{\Sigma}_\eta, \tag{2.18}$$

$$\boldsymbol{\nu}_{t|t-1}^{(i,j)} = \boldsymbol{y}_t - \boldsymbol{X}\boldsymbol{\beta}_{t|t-1}^{(i,j)}, \tag{2.19}$$

$$\boldsymbol{f}_{t|t-1} = \boldsymbol{X}\boldsymbol{P}_{t|t-1}\boldsymbol{X}' + \boldsymbol{\Sigma}_\epsilon, \tag{2.20}$$

$$\boldsymbol{\beta}_{t|t}^{(i,j)} = \boldsymbol{\beta}_{t|t-1}^{(i,j)} + \boldsymbol{P}_{t|t-1}\boldsymbol{X}'\boldsymbol{f}_{t|t-1}^{-1}\boldsymbol{\nu}_{t|t-1}^{(i,j)},$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{X}'\boldsymbol{f}_{t|t-1}^{-1}\boldsymbol{X}\boldsymbol{P}_{t|t-1}.$$

Collapsing step

Given a M-state Markov-switching component S_t, the Kalman filter produces M^2 forecasts at each time t, corresponding to every possible value for i and j. This means that, at each iteration, the number of forecasts of $\boldsymbol{\beta}_{t|t}^{(i,j)}$ increases by a factor M. Even in the MS-DNS model, where $M = 2$, this would still result in more than 1,000 forecasts to consider after just 10 iterations. Therefore, it is necessary to introduce approximations to make the above Kalman filter computationally feasible. This is done in the collapsing step, where the key idea is to collapse the $(M \times M)$ forecasts of $\boldsymbol{\beta}_{t|t}^{(i,j)}$ into M forecasts $\boldsymbol{\beta}_{t|t}^{j}$. Kim and Nelson (1999) propose the following approximations for $\boldsymbol{\beta}_{t|t}^{j}$:[14]

$$\boldsymbol{\beta}_{t|t}^{j} = \frac{\sum_{i=1}^{M} P[S_t = j, S_{t-1} = i | I_t]\boldsymbol{\beta}_{t|t}^{(i,j)}}{P[S_t = j | I_t]}. \tag{2.21}$$

At the end of each iteration, equation (2.21) is used to collapse the $(M \times M)$ forecasts of $\boldsymbol{\beta}_{t|t}^{(i,j)}$ into $(M \times 1)$ forecasts $\boldsymbol{\beta}_{t|t}^{j}$. The latter are then used as input into equation (2.17) for the next

[14]For the derivation of the approximation for $\boldsymbol{\beta}_{t|t}^{j}$, see Kim and Nelson (1999).

iteration of the Kalman filter. However, to complete the Kalman filter with the proposed approximations, one needs to make inference on the state probabilities in equation (2.21). This is done through the Hamilton filter.

Hamilton filter

At the beginning of iteration t, I start the Hamilton filter by calculating the joint probability of the states $P[S_t = j, S_{t-1} = i|I_t]$ using Bayes' rule:

$$\begin{aligned} P[S_t = j, S_{t-1} = i|I_t] &= P[S_t = j, S_{t-1} = i|I_{t-1}, \boldsymbol{y}_t] \\ &= \frac{f(\boldsymbol{y}_t, S_t = j, S_{t-1} = i|I_{t-1})}{f(\boldsymbol{y}_t|I_{t-1})} \\ &= \frac{f(\boldsymbol{y}_t|S_t = j, S_{t-1} = i, I_{t-1})P[S_t = j, S_{t-1} = i|I_{t-1}]}{f(\boldsymbol{y}_t|I_{t-1})}. \end{aligned} \tag{2.22}$$

The two terms in the numerator and the density in the denominator can be expressed in terms of known quantities. For instance, one can use the prediction error decomposition to rewrite the conditional density $f(\boldsymbol{y}_t|S_t = j, S_{t-1} = i, I_{t-1})$ as

$$f(\boldsymbol{y}_t|S_t = j, S_{t-1} = i, I_{t-1}) = (2\pi)^{-\frac{N}{2}}|\boldsymbol{f}_{t|t-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\nu}_{t|t-1}^{(i,j)\prime}\boldsymbol{f}_{t|t-1}^{-1}\boldsymbol{\nu}_{t|t-1}^{(i,j)}\right), \tag{2.23}$$

where $\boldsymbol{\nu}_{t|t-1}^{(i,j)}$ and $\boldsymbol{f}_{t|t-1}$ are obtained through the Kalman filter in equations (2.19) and (2.20), respectively. Furthermore, the second term in equation (2.22), $P[S_t = j, S_{t-1} = i|I_{t-1}]$, can be rewritten as

$$P[S_t = j, S_{t-1} = i|I_{t-1}] = P[S_t = j|S_{t-1} = i]P[S_{t-1} = i|I_{t-1}]. \tag{2.24}$$

Here, $P[S_t = j|S_{t-1} = i]$ is the transition probability and $P[S_{t-1} = i|I_{t-1}]$ is assumed to be given. The density in the denominator of equation (2.22), $f(\boldsymbol{y}_t|I_{t-1})$, is obtained by taking the sum over all states i and j:

$$\begin{aligned} f(\boldsymbol{y}_t|I_{t-1}) &= \sum_{j=1}^{M}\sum_{i=1}^{M} f(\boldsymbol{y}_t, S_t = j, S_{t-1} = i|I_{t-1}) \\ &= \sum_{j=1}^{M}\sum_{i=1}^{M} f(\boldsymbol{y}_t|S_t = j, S_{t-1} = i, I_{t-1})P[S_t = j, S_{t-1} = i|I_{t-1}]. \end{aligned} \tag{2.25}$$

Equation (2.23), (2.24) and (2.25) are then used to calculate $P[S_t = j, S_{t-1} = i|I_t]$ in equation (2.22). Finally, one needs to make inference on $P[S_t = j|I_t]$ in order to complete the collapsing step of the Kim filter. This is done by summing over all M states

$$P[S_t = j|I_t] = \sum_{i=1}^{M} P[S_t = j, S_{t-1} = i|I_t]. \tag{2.26}$$

After running the Hamilton filter, the probabilities in equation (2.24) and (2.26) are used in the collapsing step to calculate $\boldsymbol{\beta}^j_{t|t}$. Once the Kim filtering is finished, the state vector $\boldsymbol{\beta}_{t|t}$ is obtained by taking the probability-weighted average of the regime-dependent state vectors $\boldsymbol{\beta}^j_{t|t}$:

$$\boldsymbol{\beta}_{t|t} = \sum_{j=1}^{M} P[S_t = j|I_t]\boldsymbol{\beta}^j_{t|t}.$$

Maximum Likelihood

After completing the Kim filter, ML is used to obtain optimal parameter estimates. In vector notation, the $(M \times 1)$ vector of state probabilities $\pi_{t|t-1}$ and densities D_t are given by

$$\pi_{t|t-1} = \begin{pmatrix} P[S_t = 1|I_{t-1}] \\ \vdots \\ P[S_t = j|I_{t-1}] \\ \vdots \\ P[S_t = M|I_{t-1}] \end{pmatrix}, \qquad D_t = \begin{pmatrix} f(\boldsymbol{y}_t|S_t = 1, I_{t-1}) \\ \vdots \\ f(\boldsymbol{y}_t|S_t = j, I_{t-1}) \\ \vdots \\ f(\boldsymbol{y}_t|S_t = M, I_{t-1}) \end{pmatrix}. \tag{2.27}$$

The log-likelihood which the Kim filter maximizes with respect to $\boldsymbol{\theta}$ is then given as

$$\begin{aligned} \ell(\boldsymbol{\theta}) = \ln[f(\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_T)] &= \sum_{t=1}^{T} \ln[f(\boldsymbol{y}_t|I_{t-1})] \\ &= \sum_{t=1}^{T} \log[\mathbf{1}'(\pi_{t|t-1} \odot D_t)], \end{aligned} \tag{2.28}$$

where $\mathbf{1}$ is a $(M \times 1)$ vector of ones and $\odot$ denotes element-wise multiplication. The new estimates of the unknown parameters $\boldsymbol{\theta}^{(1)}$ serve as input for the first step of next iteration in the Kim filter. One obtains optimal parameter values $\boldsymbol{\theta}_{ML}$ when these parameters maximize the log-likelihood function in equation (2.28). Afterwards, the updated state probabilities $\pi_{t|t}$ are calculated as

$$\pi_{t|t} = \frac{\pi_{t|t-1} \odot D_t}{\mathbf{1}'(\pi_{t|t-1} \odot D_t)}. \tag{2.29}$$

Estimation of the three macro MS-DNS models is also done with the Kim filter and is initialised with the estimates of the two-step procedure. In the MS-DRA and MS-DRA-TVTP model, I set the mean of the k transition equations with macro-economic indicators to 0.05. For the MS-TVTP model, I initialise the $k+1$ parameters $\boldsymbol{\gamma}_j$ in the transition probabilities with zero such that the transition probability matrix P_t is initialised to be fixed. Finally, the same log-likelihood as in (2.28) is maximized in all macro models but the density of $\boldsymbol{y}_t$ now conditions on the macro-economic indicators z_{t-1} as well.

Relative to linear time series models, diagnostic checking in regime-switching models is rather difficult since the true residuals depend on the unobserved state S_t. This causes the true residuals to be unobserved as well. To compare the in-sample performance of the DNS model and regime-switching models, I therefore follow Smith (2008) and use the Rosenblatt (1952) transformation. Let μ_j and σ_j^2 be the conditional mean and conditional variance of $\boldsymbol{y}_t$ in regime j, respectively. Then the Rosenblatt residual $\tilde{e}_t$ is defined as

$$\tilde{e}_t = \Phi^{-1}\left(\sum_{j=1}^{M} P[S_t = j | I_{t-1}]\Phi(\sigma_j^{-1}(y_t - \mu_j))\right), \tag{2.30}$$

where Φ^{-1} is the inverse of a standard normal cumulative distribution function. If $\boldsymbol{y}_t$ is generated by the distribution implied by the model, then $\tilde{e}_t$ is i.i.d. and standard normally distributed. In calculating the Rosenblatt residuals, I identify the yield to be in regime j when the filtered probability for regime j is greater than 0.5.

2.3 Likelihood Ratio Test

Since all discussed models are extensions of the baseline DNS model, I assess the significance of the extended models by performing likelihood ratio (LR) tests. For instance, I employ this test to evaluate the additional value of the macro-economic indicators in the three macro regime-switching models relative to the MS-DNS model.

Statistically testing for the presence of a second regime is, however, not as straightforward. This is because the MS-DNS model is not fully nested within the DNS model due to the imposed restriction that f_{22} is equal to zero. Moreover, under the null hypothesis of a single regime, the transition probabilities and regime-switching parameter $\mu_{2,2}$ are not identified. This causes the LR test statistic no longer to be approximately χ^2 distributed such that LR test results are incorrect. Consequentially, the three macro regime-switching models are also not nested within the DNS model. An overview of which models are nested within the baseline DNS and MS-DNS model is provided in Table 1. Here, I note that I slightly modify the DRA model and set f_{22} equal to zero to keep the model specification consistent with the (macro) regime-switching

Table 1: Overview nested models within baseline DNS and MS-DNS model

	DNS	DRA	MS-DNS	MS-DRA	MS-TVTP	MS-DRA-TVTP
DNS		×	×	×	×	×
MS-DNS				✓	✓	✓

Note: This table gives an overview of which models are nested within the baseline DNS and MS-DNS model. A checkmark (✓) means that the considered models are indeed nested. A cross (×) indicates that it statistically incorrect to perform the likelihood ratio test since the considered models are not fully nested.

models. This causes the DRA and DNS model not to be nested anymore.

Multiple alternatives exist to overcome the problem of a non-χ^2 distributed LR statistic when testing for the existence of multiple regimes. Hansen (1996) introduces a standardized LR test for nonlinear models such as regime-switching models. In this test, the identified parameters are concentrated out of the likelihood function (also called the 'concentrated' likelihood) which is then maximized with respect to the unidentified parameters through a grid search. However, Garcia (1998) shows that the standardized LR test of Hansen (1996) is computationally demanding, even when the amount of model parameters is quite small.

Therefore, numerous studies still resort to the LR test to provide a certain degree of confidence about the significance of a second regime. For instance, Gelman and Wilfling (2009) estimate a Markov-switching generalized autoregressive conditional heteroscedasticity (GARCH) model and use bootstrapping techniques to evaluate the finite-sample distribution of their LR test statistic. They report the finite distribution to be comparable to a χ^2 distribution with the degrees of freedom equal to the number of parameter restrictions. Due to this promising result and because existing tests for regime-switching models require substantial computer power, I also recede to the LR test.

3 Data

This section discusses the data used in this thesis. Section 3.1 studies the US Treasury yield data. In Section 3.2, I describe the macro-economic indicators which are used to extend the MS-DNS model and elaborate on the relation between these indicators and the yield curve.

3.1 US Treasury Yields

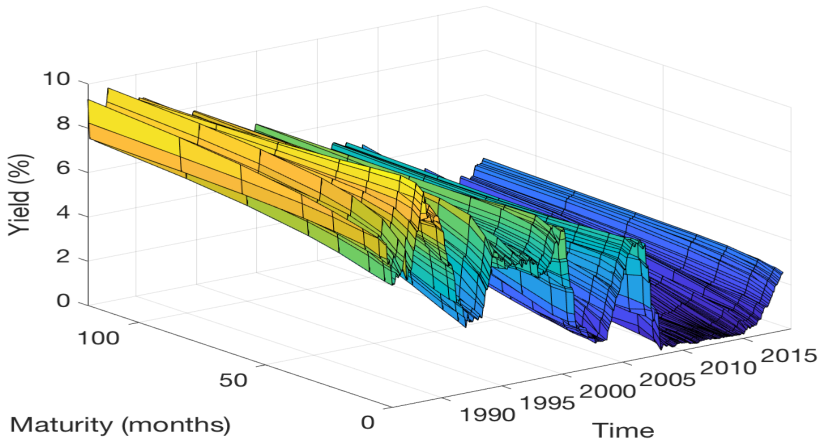
I use end-of-the-month US yield data from January 1986 to December 2018. For the short end of the yield curve, I use three- and six-month Treasury yields from the H.15 database of the Fed. This yield data is combined with continuously compounded smoothed zero-coupon Treasury yields with maturities of one, two, three, five, seven and ten years from the Gürkaynak et al. (2007) database.[15] This database provides Treasury yield curve estimates based on the Svensson (1994) model which is an extension of the Nelson-Siegel model by including a second curvature factor. Combining yield data from these two databases is done before by studies such as Christensen (2013) and Bauer and Rudebusch (2016).

An important reason to particularly study this sample period is due to the recent periods of extreme low interest rates in the US. Gaussian term structure models, including the Nelson-Siegel model, typically do not incorporate the possibility of interest rates to reach the ZLB since this rarely happens in practice. Therefore, Gaussian type of models often provide a poor approximation of the yield curve near the ZLB.

Figure 2 presents a surface plot of the yield curve against maturity and time. It shows that yields vary significantly over time and that interest rates are extremely low from the start of the global financial crisis in 2008. Furthermore, it is evident that the yield curve is not constant over time but can take a variety of shapes. During the beginning of the sample period, the yield curve is nearly flat, whereas it is steeply upward sloping during the global financial crisis.

[15] The yield data sets are available at https://www.federalreserve.gov/releases/h15/ and https://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

Figure 2: Yield curve surface plot



Note: This figure shows a surface plot of the monthly US yield data, spanning the period January 1986 to December 2018 at maturities of 3 and 6 months, and 1, 2, 3, 5, 7, and 10 years.

Table 2 presents summary statistics of the US Treasury yields data and of the empirical level, slope and curvature factors which are similarly defined as in Diebold and Li (2006). The empirical level factor is equal to the 120-month yield. The empirical slope factor is the spread of the 3-month yield over the 120-month yield, and the empirical curvature factor is twice the 24-month yield minus the sum of the 3-month and 120-month yield.

The statistics show that, in general, the yield curve is upward sloping, and that the long end is less volatile and more persistent over time than the short end. However, the 3-month Treasury yield is still quite persistent with an autocorrelation of 0.519 after 30 months. Furthermore, the level factor is highly persistent and has a relatively small standard deviation. On the other hand, the slope factor is the least persistent factor and highly variable around its mean. This is in contrast to previous literature which often finds the curvature factor to be the least persistent. My finding regarding the slope factor is most likely due to the change from a nearly flat to a steep curve during the end of the sample period where interest rates are extremely low.[16]

It is known from the literature that the first three principal components summarize most of the information contained in the yield curve (see Litterman and Scheinkman (1991)). In Table C.12 of the Appendix, I perform a principal component analysis and indeed find that the first three components account for 99.97% of the total variation in the yield curve. The first component explains 97.13% of the variation in the data and has a positive loading across all maturities. Hence, the first component may be interpreted as a level factor as a shock would cause all yields to change in the same direction. The second component captures 2.66% of the

[16] Levant and Ma (2017) also find the slope factor to be the least persistent factor in periods of low interest rates due to the yield curve being flat.

Table 2: Summary statistics Treasury yields (in %)

Maturity	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}_1$	$\hat{\rho}_{12}$	$\hat{\rho}_{30}$
3	3.285	2.573	0.000	9.230	0.993	0.845	0.519
6	3.432	2.613	0.030	9.540	0.993	0.850	0.528
12	3.617	2.635	0.099	9.658	0.993	0.863	0.567
24	3.874	2.611	0.188	9.566	0.991	0.877	0.633
36	4.108	2.543	0.306	9.459	0.990	0.885	0.679
60	4.511	2.398	0.627	9.317	0.988	0.891	0.730
84	4.837	2.285	1.007	9.406	0.987	0.892	0.751
120 (level)	5.195	2.174	1.498	9.642	0.986	0.890	0.760
Slope	1.910	1.205	−0.616	4.376	0.966	0.442	−0.286
Curvature	−0.731	0.900	−2.691	1.634	0.949	0.596	0.148

Note: This table presents summary statistics of the US Treasury yields, spanning the period January 1986 to December 2018, measured in percentages on a monthly basis. For each maturity, I show the mean, standard deviation, minimum, maximum and the autocorrelation coefficient at 1 month ($\hat{\rho}_1$), 1 year ($\hat{\rho}_{12}$), and 30 months ($\hat{\rho}_{30}$).

variation in the Treasury yields and has considerable negative loadings at the short end and substantial positive loadings at the long end of the yield curve. Therefore, this component serves as a slope factor since shocks to this factor determine whether the yield curve flattens or becomes steeper. Lastly, the third component, which accounts for 0.18% of the yield variation, is hump shaped as the loadings are only negative for the middle part of the curve and can thus be interpreted as a curvature factor. This motivates the use of the Nelson-Siegel model, which also uses a level, slope and curvature factor, for modelling the Treasury yields.[17]

3.2 Macro-Economic Indicators

As macro-economic indicators, I consider the real gross domestic product (GDP) growth, the Federal funds rate (FFR), and the percent change in the seasonally-adjusted consumer price index (CPI), which are all obtained from the database of the Federal Reserve Bank of St. Louis (FRED).[18] I consider these three indicators as they are widely considered to be the minimum set of economic fundamentals required to capture macro-economic dynamics (see Rudebusch and Svensson (1999) and Kozicki and Tinsley (2001)). The three variables namely represent the level of real economic activity relative to potential, the monetary policy instrument, and the inflation rate, respectively. GDP (CPI) growth is measured as the percentage change relative to the same quarter (month) one year ago (also known as year-over-year changes). Furthermore, as GDP is only available at a quarterly frequency, I assume that GDP growth is equal for the months within each quarter.

Panel (A) of Table 3 presents summary statistics of the macro-economic indicators. The

[17] It should be noted, however, that the principal component factors are not identical to the estimated level, slope and curvature factor from the Nelson-Siegel model.

[18] Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/.

year-over-year changes of both GDP and CPI are on average positive. The average FFR is far away from zero but quite volatile around its mean. Furthermore, both GDP and CPI growth attain their minimum value around the global financial crisis in 2008. It was also during this period that the Fed started its first round of quantitative easing to fight domestic deflation. The American central bank had to rely on such an unconventional form of monetary policy since its policy rate was already near the ZLB. This is also reflected in the minimum FFR of seven basis points which is reached approximately two years after the crisis.

Table 3: Summary statistics macro-economic indicators

Panel (A): Summary statistics (in %) and autocorrelation coefficients

Macro Factor	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}_1$	$\hat{\rho}_{12}$	$\hat{\rho}_{30}$
GDP Growth	2.618	1.636	-3.900	5.300	0.963	0.331	0.032
FFR	3.492	2.744	0.070	9.850	0.994	0.842	0.507
CPI Growth	2.611	1.333	-2.097	6.290	0.957	0.294	0.149

Panel (B): Correlation with empirical factors and first three principal components

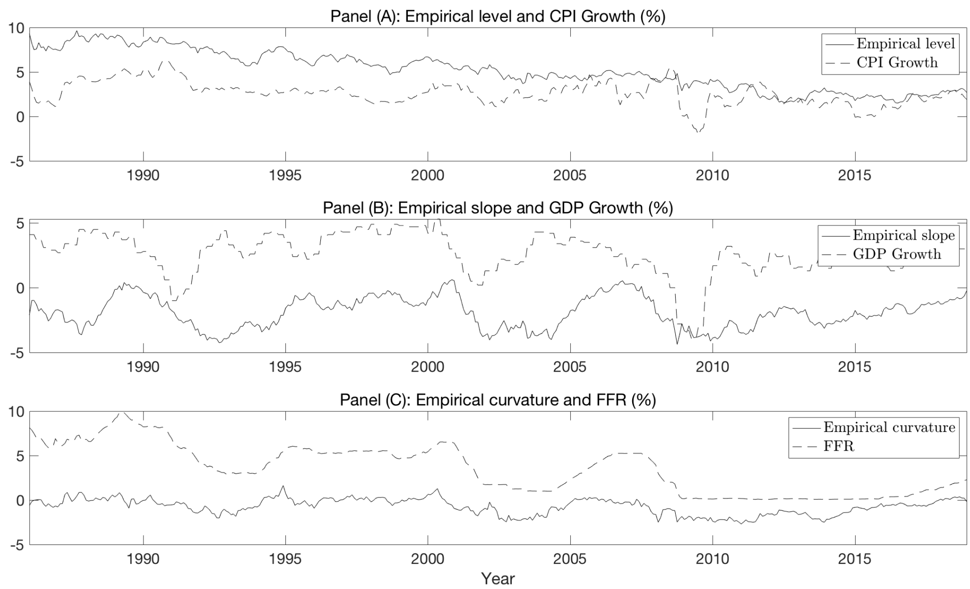
Macro Factor	Level	Slope	Curvature	First P.C.	Second P.C.	Third P.C.
GDP Growth	0.310	-0.307	0.523	0.398	-0.177	-0.300
FFR	0.879	-0.537	0.679	0.972	-0.187	0.098
CPI Growth	0.588	-0.260	0.346	0.616	-0.043	0.162

Note: This table presents summary statistics of the macro-economic indicators, spanning the period January 1986 to December 2018. In Panel (A), I show the mean, standard deviation, minimum, maximum (all in %) and the autocorrelation coefficient at 1 month ($\hat{\rho}_1$), 1 year ($\hat{\rho}_{12}$), and 30 months ($\hat{\rho}_{30}$). Panel (B) presents correlations between the indicators, the empirical factors and the first three principal components. The latter are obtained from the principal component analysis shown in Table C.12 of the Appendix.

In Figure 3, I plot time series of the empirical factors and the three macro-economic indicators. The three indicators follow roughly the same pattern as they are all quite stable before the start of the global financial crisis and drop dramatically during the crisis. From the plots, one may see a clear relation between GDP growth, the FFR and the inflation rate (measured by CPI). During times of economic contraction, unemployment rates often increase while wages decrease. This results in a decrease of the inflation rate. The Fed tries to prevent the inflation from falling by stimulating the economy. One way is by lowering the FFR which makes borrowing money cheaper. Therefore, consumers have more money to spend, causing GDP to grow and inflation to increase. The pattern described above is clearly visible during, for example, the period of 2001 and onward.

Figure 3 also shows that the empirical yield factors are closely linked to the macro-economic indicators. This is confirmed by the correlations between the indicators and empirical factors, which are shown in the first three columns of Panel (B) of Table 3. The correlation between the empirical level and CPI growth is 0.588; for the empirical slope and GDP growth this is 0.307. The magnitude of these correlations is consistent with those found in DRA. Furthermore, the correlation between the empirical curvature and FFR is 0.679. The last three columns of Panel (B) of Table 3 also show that the correlations between the indicators and the first three principal components are of a similar size (except for the correlation between GDP growth and the third component). This supports the interpretation of the first three components as a level, slope and curvature factor. Since CPI, GDP growth and the FFR are closely linked to the first three components and the empirical level, slope and curvature, it is sensible to include these three macro-economic indicators in the regime-switching models.

Figure 3: Time series of empirical factors and macro-economic indicators



Note: This figure shows time series of the empirical level, slope and curvature factors together with three closely linked macro-economic indicators, spanning the period January 1986 to December 2018. GDP growth and CPI growth are measured based on year-over-year changes.

4 In-sample Results

In this section, I present in-sample results of the various Nelson-Siegel models. Section 4.1 describes the findings on the baseline DNS model. Section 4.2 presents the model of DRA which extends the DNS model with macro-economic indicators. In Section 4.3, I discuss the in-sample fit of the MS-DNS model which allows for regime-switching in the baseline model. Extensions of the MS-DNS model with macro-economic indicators are discussed in Section 4.4 - 4.6.

4.1 DNS: Baseline Dynamic Nelson-Siegel Model

In Panel (A) of Table 4, I report estimation results for the baseline DNS model where the latent factors follow a restricted AR(1) process, as shown in equations (2.6) and (2.7) of Section 2.1.2. Estimates of the autoregressive matrix $\boldsymbol{F}$ show that all three Nelson-Siegel factors are significant and highly persistent. Furthermore, the level is the most persistent and the curvature the least persistent factor. This is in line with the first-order autocorrelation coefficients of the empirical factors found in Table 2 of Section 3.1. The mean level and slope are both insignificantly different from zero; only the mean curvature is significant at the five percent level. Estimates of the covariance matrix $\boldsymbol{\Sigma}_\eta$ are presented in Panel (B). The volatility of the transition disturbances increases when we move from the level to the slope to the curvature factor. Finally, the estimated decay parameter λ of 0.040 is highly significant and implies that the loading on the curvature factor attains its maximum at a maturity of 44.8 months.

Table 4: Estimates of the DNS model - AR specification

Panel (A): Autoregressive matrix $\boldsymbol{F}$ and vector of means $\boldsymbol{\mu}$

	Level_{t-1}	Slope_{t-1}	Curvature_{t-1}	$\boldsymbol{\mu}$
Level_t ($\beta_{1,t}$)	**0.993**			0.041
	(0.005)			(0.030)
Slope_t ($\beta_{2,t}$)		**0.986**		-0.032
		(0.008)		(0.025)
Curvature_t ($\beta_{3,t}$)			**0.942**	**–0.088**
			(0.017)	(0.048)

Panel (B): Covariance matrix $\boldsymbol{\Sigma}_\eta$ and decay parameter λ

	Level_t ($\beta_{1,t}$)	Slope_t ($\beta_{2,t}$)	Curvature_t ($\beta_{3,t}$)	λ
Level_t ($\beta_{1,t}$)	**0.066**			**0.040**
	(0.004)			(0.000)
Slope_t ($\beta_{2,t}$)		**0.112**		
		(0.007)		
Curvature_t ($\beta_{3,t}$)			**0.942**	
			(0.060)	

Note: This table reports estimates of the baseline DNS model where the latent factors follow an AR(1) process. Panel (A) presents estimates of the autoregressive coefficient matrix $\boldsymbol{F}$ and vector of means $\boldsymbol{\mu}$. Panel (B) shows estimates of the covariance matrix $\boldsymbol{\Sigma}_\eta$ and decay parameter λ. Bold entries denote parameter estimates significant at the five percent level. Standard errors appear in parentheses.

For completeness, I also present parameter estimates of the DNS model where the latent factors follow a VAR(1) process. Table C.13 of the Appendix shows that all estimated coefficients and standard errors are of equal magnitude relative to the restricted AR(1) DNS model. Moreover, only one of the off-diagonal elements in $\boldsymbol{F}$ and none of the off-diagonal elements in $\boldsymbol{\Sigma}_\eta$ are significant. Since both versions of the DNS model give similar parameter estimates in-sample but the parsimonious AR(1) version has found to produce significantly better forecasts out-of-sample (see Christensen et al. (2011)), I use the restricted AR(1) DNS model as the baseline model for its extensions.

To evaluate the in-sample performance of the Nelson-Siegel models, the first column of Table 5 reports residual diagnostics of the DNS model based on the Rosenblatt transformation as shown in equation (2.30) of Section 2.2.3. According to the Jarque-Bera test statistics in Panel (A), normality is rejected for 7 out of the 8 maturities at the five percent level. This suggests that the DNS model does not fit the yield data quite well. The Ljung-Box test statistics for the residuals and squared residuals in Panel (B) and (C) do, however, indicate that the Rosenblatt residuals are i.i.d. distributed.

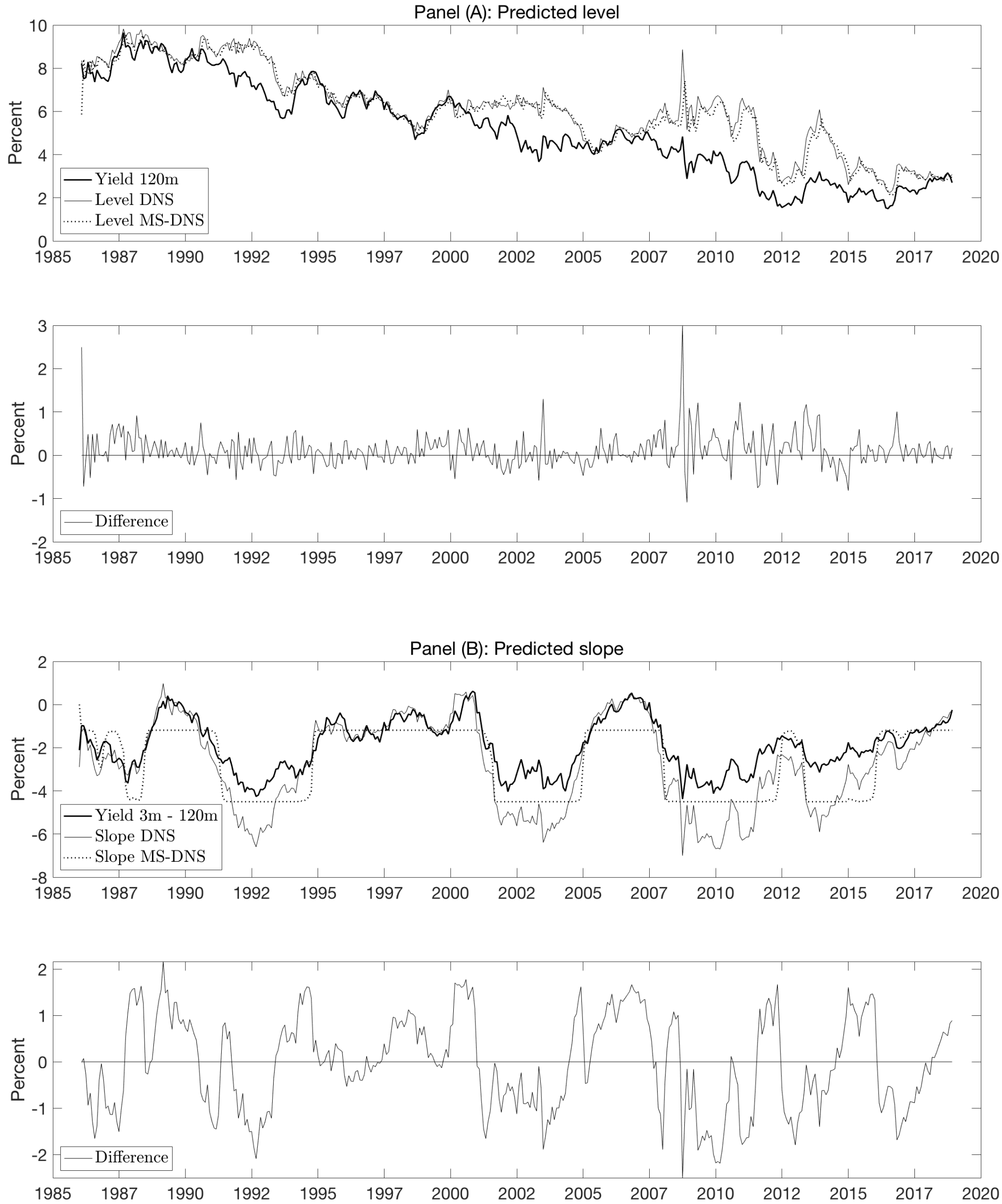
I use the Kalman filter to obtain estimates of the latent Nelson-Siegel factors. The top panel of Figure 4 shows the predicted latent factors together with their empirical proxies. The level factor is positive but steadily decreases over time. In contrast, the slope and curvature factor are mostly negative. The loading on the slope factor also switches between a state where it is negative and one where its loading is near zero. Finally, all three factors are closely linked to their empirical proxies with a correlation of 0.82, -0.85 and 0.78, respectively, which supports the interpretation of the latent factors as a level, slope and curvature factor. Only starting from the beginning of the global financial crisis in 2008, the DNS model tends to produce more extreme estimates of the latent factors. That is, the estimated factor is more positive when the empirical proxy is positive, and more negative when the empirical proxy is negative.

Table 5: Rosenblatt residual diagnostics

Maturity (months)	DNS	DRA	MS-DNS	MS-DRA	MS-TVTP	MS-DRA-TVTP
Panel (A): Jarque-Bera test						
3	0.028*	0.047*	0.050*	0.025*	0.010*	0.124
6	0.045*	0.071	0.107	0.111	0.141	0.119
12	0.042*	0.068	0.105	0.111	0.072	0.064
24	0.055	0.041*	0.067	0.107	0.118	0.126
36	0.040*	0.031*	0.094	0.028*	0.062	0.104
60	0.038*	0.061	0.089	0.062	0.108	0.103
84	0.034*	0.029*	0.081	0.022*	0.106	0.100
120	0.030*	0.057	0.037*	0.087	0.105	0.129
Panel (B): Ljung-Box test						
3	0.409	0.386	0.730	0.779	0.810	0.795
6	0.421	0.522	0.737	0.785	0.824	0.810
12	0.395	0.478	0.701	0.758	0.800	0.778
24	0.390	0.355	0.698	0.741	0.789	0.772
36	0.378	0.336	0.684	0.729	0.772	0.763
60	0.381	0.444	0.715	0.740	0.794	0.791
84	0.412	0.361	0.733	0.761	0.810	0.804
120	0.347	0.401	0.589	0.678	0.718	0.689
Panel (C): Ljung-Box test squared residuals						
3	0.346	0.315	0.518	0.576	0.612	0.600
6	0.381	0.402	0.555	0.594	0.634	0.618
12	0.339	0.359	0.514	0.558	0.601	0.584
24	0.335	0.308	0.512	0.547	0.594	0.571
36	0.333	0.295	0.483	0.539	0.589	0.567
60	0.335	0.356	0.494	0.545	0.595	0.580
84	0.369	0.310	0.491	0.573	0.608	0.581
120	0.321	0.348	0.427	0.518	0.560	0.522

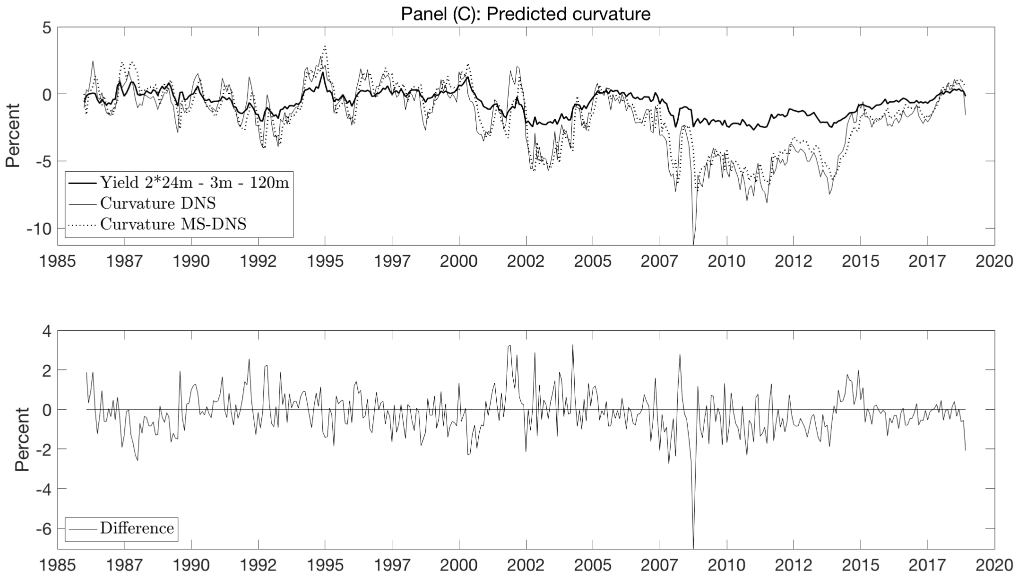
Note: This table reports residual diagnostics of the Nelson-Siegel models based on the Rosenblatt transformation, as shown in equation (2.30) of Section 2.2.3. Panel (A), (B) and (C) show p-values of the Jarque-Bera normality test, Ljung-Box serial correlation test, and the Ljung-Box test for squared residuals, respectively. An asterisk (*) denotes significance at the five percent level.

Figure 4: Predicted Nelson-Siegel factors of the DNS and MS-DNS model



Note: This figure presents the predicted level, slope, and curvature obtained from the DNS model (bold line) and the MS-DNS model (dashed line). Panels (A), (B), and (C) show the predicted level, slope, and curvature factor, respectively, of both models together with their empirical proxies. The empirical level is equal to the 120-month yield. The slope is the spread of the 3-month yield over the 120-month yield, and the curvature is twice the 24-month yield minus the sum of the 3-month and 120-month yield. In the bottom plot of each panel, I present the differences of the predicted level, slope and curvature estimates of the DNS and MS-DNS model.

Figure 4: (Continued.)



4.2 DRA: The DNS Model with Macro-Economic State Variables

In this section, I follow the work of DRA and include macro-economic indicators as state variables in the baseline DNS model. Recall that I set f_{22} to zero to keep the model specification consistent with the specification of the (macro) regime-switching models. In addition, I only let the upper right (3×3) block unrestricted in the autoregressive matrix $\boldsymbol{F}^*$ such that there is a macro-to-yields interaction but the yields do not give any feedback to the indicators. This is, again, consistent with the (macro) regime-switching models.

The parameter estimates of the DRA model in Table 6 are quite similar to those found by DRA. First, the level, curvature and the macro-economic indicators are all highly significant and persistent. Second, GDP growth is a significant predictor for the slope factor whereas the Federal funds rate significantly influences both the level and slope factor.[19]

Residual diagnostics of the DRA model, presented in the second column of Table 5 in Section 4.1, show that the inclusion of macro-economic indicators in the baseline DNS model marginally improves the in-sample fit. The p-values of the Jarque-Bera and Ljung-Box statistics slightly increase for 4 out of the 8 maturities. The minor improvement of the DRA model is also profound in terms of log-likelihood and information criteria values, which are shown in Table 7. The addition of 18 parameters in the DRA model improves the log-likelihood value by just 118.0. The decrease in the Akaike information criterion (AIC) and the Bayesian information

[19]The only difference between the indicators used in the original model of DRA and in this thesis is that DRA use manufacturing capacity utilization instead of GDP growth as a proxy for the level of real economic activity relative to potential.

criterion (BIC) value is of a similar magnitude. However, the value of the LR test statistic of 236.0 gives an indication that the performance of the DRA model over the baseline model is still sizeable.

Table 6: Estimates of the DRA model

	Level_{t-1}	Slope_{t-1}	Curvature_{t-1}	GDP_{t-1}	FFR_{t-1}	CPI_{t-1}	$\boldsymbol{\mu}$	λ
Level_t $(\beta_{1,t})$	**0.951**			-0.018	**0.024**	-0.006	**1.327**	**0.039**
	(0.032)			(0.021)	(0.028)	(0.012)	(1.120)	(0.004)
Slope_t $(\beta_{2,t})$				**0.036**	**0.341**	0.031	**−0.518**	
				(0.015)	(0.071)	(0.028)	(0.367)	
Curvature_t $(\beta_{3,t})$			**0.878**	0.042	0.008	-0.002	-0.412	
			(0.031)	(0.017)	(0.074)	(0.043)	(0.437)	
GDP_t				**0.998**			**1.586**	
				(0.012)			(0.452)	
FFR_t					**0.991**		**0.741**	
					(0.037)		(0.439)	
CPI_t						**0.976**	**0.418**	
						(0.010)	(0.448)	

Note: This table reports a selection of the parameter estimates of the DRA model. The table presents estimates of the autoregressive coefficient matrix $\boldsymbol{F}^*$, vector of means $\boldsymbol{\mu}$ and decay parameter λ. Bold entries denote parameter estimates significant at the five percent level. Standard errors appear in parentheses.

Table 7: Log-likelihood and information criteria of the Nelson-Siegel models

	Log-likelihood	# parameters	AIC	BIC	LR-stat. (DNS)	LR-stat. (MS-DNS)
DNS	4240.0	18	−8443.9	−8372.2		
DRA	4358.0	36	−8644.0	−8500.7	236.0(×)	
MS-DNS	4930.8	20	−9821.6	−9741.9	1381.7(×)	
MS-DRA	5067.1	38	−10 058.3	−9907.0	1654.2(×)	272.7
MS-TVTP	5012.4	26	−9972.9	−9869.3	1544.8(×)	163.3
MS-DRA-TVTP	5148.4	44	−10 208.8	−10 033.6	1816.8(×)	435.2

Note: This table presents log-likelihood values, the number of parameters (# parameters), the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the DNS model and its extensions with regime-switching and macro-economic indicators. In the last two columns, I present likelihood ratio test statistics (LR-stat.) of the extended models against the baseline DNS model, and of the regime-switching macro models against the MS-DNS model, respectively. A cross (×) indicates that it is statistically incorrect to perform the likelihood ratio test since the considered models are not fully nested. In these cases, the likelihood ratio test statistic is reported to provide a certain degree of confidence about the improvement of the considered model.

4.3 MS-DNS: The DNS Model with Markov-Switching

The second modification of the baseline DNS model allows for regime-switching in the mean of the latent slope factor. Panel (A) of Table 8 presents estimation results for the MS-DNS model, which modifies the transition equation of the DNS model as given in equation (2.11) of Section 2.2.1. The switching mean has two distinct states that are both highly significant. In the first state, $\mu_{2,1}$ equals -1.190 indicating a nearly flat yield curve and, hence, a period where short and long term interest rates do not differ substantially. The second state corresponds to a steep yield curve as $\mu_{2,2}$ equals -4.503. The two estimates of the regime-switching mean accurately coincide with the two states found in the empirical slope in Figure 4 of Section 4.1. The estimated transition probabilities in Panel (B) show that both states are highly persistent. This is economically very encouraging since it indicates that within each regime the shape of the term structure is relatively stable across time. Contrary to the DNS model, the volatility in the transition disturbances of $\mathbf{\Sigma}_\eta$ is the highest for the slope factor. This is due to the specification of the MS-DNS model which sets the middle element of $\boldsymbol{F}$, f_{22}, equal to zero. Hence, all the variability in the slope factor is accounted for in the regime-switching mean and the transition error.

Table 8: Estimates of the MS-DNS model

Panel (A): Autoregressive matrix $\boldsymbol{F}$, vector of means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and decay parameter λ

	Level_{t-1}	Slope_{t-1}	Curvature_{t-1}	$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_2$	λ
Level_t ($\beta_{1,t}$)	**0.991**			0.052		**0.041**
	(0.006)			(0.036)		(0.001)
Slope_t ($\beta_{2,t}$)				**-1.190**	**-4.503**	
				(0.089)	(0.107)	
Curvature_t ($\beta_{3,t}$)			**0.944**	-0.086		
			(0.017)	(0.047)		

Panel (B): Covariance matrix $\mathbf{\Sigma}_\eta$, and transition probabilities p_{11} and p_{22}

	Level_t ($\beta_{1,t}$)	Slope_t ($\beta_{2,t}$)	Curvature_t ($\beta_{3,t}$)	p_{11}	p_{22}
Level_t ($\beta_{1,t}$)	**0.082**			**0.975**	**0.971**
	(0.007)			(0.011)	(0.012)
Slope_t ($\beta_{2,t}$)		**1.038**			
		(0.078)			
Curvature_t ($\beta_{3,t}$)			**0.622**		
			(0.066)		

Note: This table reports estimates of the MS-DNS model. Panel (A) presents estimates of the autoregressive coefficient matrix $\boldsymbol{F}$, vector of means during periods of normal interest rates ($\boldsymbol{\mu}_1$) and low interest rates ($\boldsymbol{\mu}_2$), and decay parameter λ. Panel (B) presents estimates of the covariance matrix $\mathbf{\Sigma}_\eta$, and the transition probabilities p_{11} and p_{22}. Bold entries denote parameter estimates significant at the five percent level. Standard errors appear in parentheses.

The third column of Table 5 in Section 4.1 shows residual diagnostics of the MS-DNS model. It shows that the model fits the term structure considerably better than the baseline DNS model as normality is only rejected for 2 out of the 8 maturities. The Ljung-Box statistics for the MS-DNS model are also considerably larger. Similar to the DNS model, the 3-month and 120-month yield are the most difficult to fit for the regime-switching model indicated by the significant p-value of the Jarque-Bera test. However, the crucial middle part with maturities ranging from 6 to 60 months is fit substantially better as 4 out of the 6 improvements of in-sample fit occur in the middle part of the curve. Table 7 of Section 4.2 presents the performance of the Nelson-Siegel models in terms of their log-likelihood, AIC and BIC values, as well as LR test statistics. The difference in log-likelihood between the MS-DNS and DNS model of 690 is very encouraging. This is also shown by the substantial decrease in AIC and BIC values. In addition, the LR statistic of 1381.7 provides strong evidence for the presence of the second regime.

In the top panel of Figure 4 in Section 4.1, I plot the predicted latent factors of the MS-DNS model. The predicted level and curvature obtained from the DNS and MS-DNS model do not deviate substantially from their empirical proxies until the start of 2008 and onward. The difference between the predicted slope of both models is more apparent. The slope of the MS-DNS model is indeed able to distinguish between the two states, which is in line with the parameter estimates found for the regime-switching mean in Table 8. In addition, when the empirical slope is near zero, the MS-DNS model slightly underestimates the slope relative to the DNS model. However, the predicted slope of the regime-switching model follows its empirical proxy considerably better when the empirical slope is far away from zero.

The bottom panel of Figure 4 presents the difference between the predicted factors of both models. A difference above (below) the zero line indicates an overestimation (underestimation) of the DNS model relative to the MS-DNS model. The difference is mostly visible for the slope and curvature factor, especially during economic recessions such as the Mexican peso crisis in 1994, the dot-com bubble in 2002, and the global financial crisis in 2008. For the level factor, the difference between the two models is less apparent with only a major outlier in 2008. The predicted factors obtained from the MS-DNS model follow their empirical proxies slightly better during recessions. For instance, in 2008 the DNS model overestimates the level factor relative to the MS-DNS model. This indicates that the filtered level of the MS-DNS model is closer to its empirical proxy (as the empirical level is lower than the filtered level of both models). The finding above is also confirmed in terms of correlation. The inclusion of regime-switching improves the correlation between the latent and empirical factors from 0.82, -0.85 and 0.78 to 0.93, -0.96 and 0.95 for the level, slope and curvature, respectively.

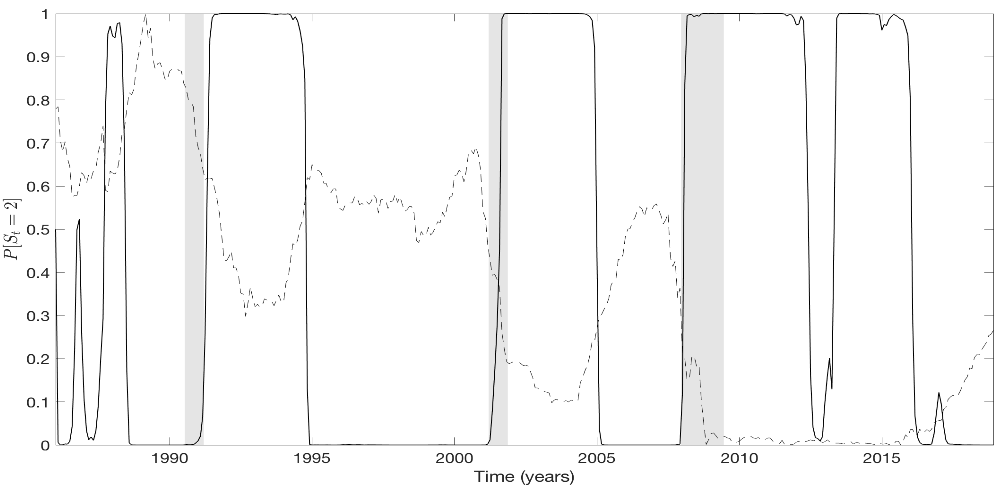
I present filtered probabilities of the low interest rate regime (regime 2) obtained from the Kim filter in Figure 5. It confirms that the second regime is very persistent since regime switches only occur after an extended time period. This behaviour of the term structure in both regimes naturally links to the evolution of the business cycle.[20] Next to the filtered probabilities, I show NBER recession periods which include economic crises such as the early 1990s recession, the dot-com bubble, and the global financial crisis. The NBER recessions coincide fairly well with the filtered probabilities and the MS-DNS model is able to identify the start of all three major crises during the sample period.

Figure 5 also displays the short-term interest rate, taken to be the 3-month yield, which is normalized between 0 and 1. It shows that periods where the filtered probability for the low interest rate regime is close to one correspond to periods where the short rate decreases substantially. Furthermore, the short rate remains relatively low during the low interest rate regime until the filtered probability decreases towards zero (e.g. during the periods 1990 - 1995, 2000 - 2005, and 2008 - 2016). Particularly interesting is the period after 2010 which the MS-DNS model recognizes as a low interest rate regime, even though it does not coincide with a recession. This indicates that the MS-DNS model is not only able to recognize periods of low interest rates due to the start of an economic recession but also when the economy is relatively stable.

In Figure 6, I assess the in-sample fit of the MS-DNS model in both regimes as well as over the full sample. The estimated yields in each regime are obtained by including the yield in regime j when the filtered probability for regime j is greater than 0.5, for each month t. Figure 6 confirms that the first regime corresponds to a nearly flat curve whereas the second regime exhibits a steep curve. Furthermore, the fitted short rate is, on average, 2.96% lower in the second regime. This indeed indicates that the first regime resembles a period with normal interest rate levels, and the second regime a period with low interest rates. Consequently, the average yield curve over the full sample falls between the two regime-dependent curves. For completeness, I also compare the fitted curves in each regime by taking the probability-weighted average between the fitted yields and the filtered state probabilities. These are shown in Figure C.13 of the Appendix. The shape of the fitted curves is similar to those presented in Figure 6 and the average fitted short rate in the second regime is also substantially lower than in the first regime.

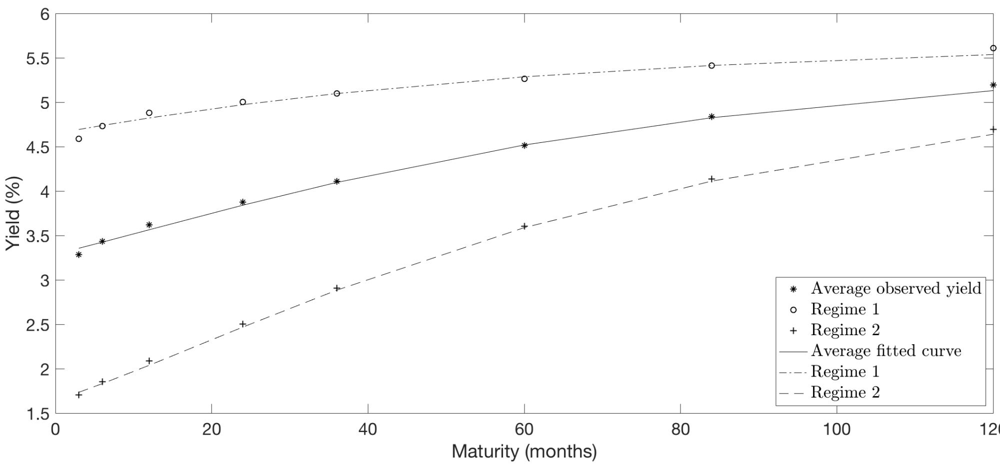
[20] I also estimate a version of the MS-DNS model which relaxes the assumption that f_{22} equals zero and sets $\boldsymbol{F}$ to a diagonal matrix. Figure C.12 of the Appendix presents the filtered probabilities of this model specification. It shows that regime-switching occurs less frequently and the two states are less persistent which is intuitively unappealing. This confirms the claim of Bernadell et al. (2005) that setting f_{22} to zero in the MS-DNS model is preferred.

Figure 5: Filtered probabilities second regime of the MS-DNS model



Note: This figure shows filtered probabilities for the second regime, a regime with low interest rates, based on the MS-DNS model. It also plots the 3-month yield (dashed line), normalized between 0 and 1. Shaded areas correspond to NBER recession periods.

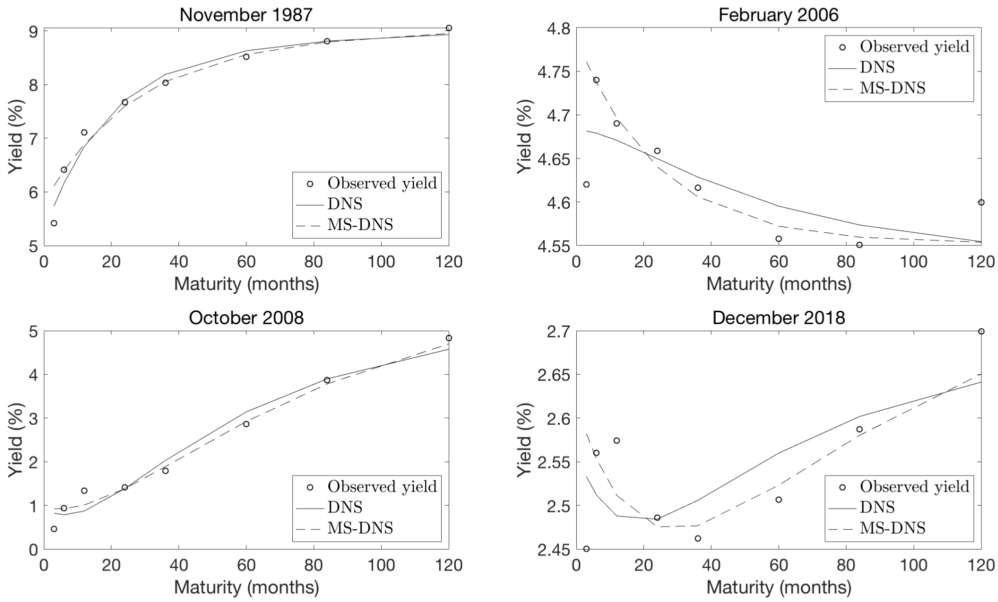
Figure 6: Average fit of the MS-DNS model



Note: This figure shows the average fit of the MS-DNS model. I show the average fitted yield curve over the full sample (solid line), and in each regime (dashed line). The regime-dependent curves are obtained by including the yield in regime j when the filtered probability for regime j is greater than 0.5. Observed yields are presented as dots, stars and crosses.

The difference in in-sample fit between the DNS and MS-DNS model becomes more visible when one considers the fit for different yield curve shapes. Figure 7 reveals that both models fit the curve well when it is upward sloping (e.g. November 1987). However, the small hump in the short end of the curve in October 2008 is not fit by both models. The fitting capability of the MS-DNS model is more apparent for hump shaped curves such as in February 2006 and December 2018. Both curves are fit remarkably well by the MS-DNS model, except for the 3-month and 120-month yield, which is consistent with the residual diagnostics found in Table 5 of Section 4.1. Interestingly, the DNS model seems to take the relatively low 3-month yield into account by underestimating the entire short end of the yield curve.

Figure 7: Fitted yield curve of the DNS and MS-DNS model for various shapes



Note: This figure shows the fitted yield curve of the DNS (solid line) and MS-DNS model (dashed line) for various yield curve shapes. Observed yields are presented as dots.

4.4 MS-DRA: Macro-Economic State Variables

Next, I introduce the addition of macro-economic indicators as state variables in the MS-DNS model. Table 9 presents a selection of the parameter estimates of the MS-DRA model, shown in equations (2.13) and (2.14) of Section 2.2.2. The parameter estimates of the MS-DRA model are very similar to those of the DRA model discussed in Section 4.2. That is, the level and curvature factor as well as the macro-economic indicators are all highly persistent and significant.[21] Furthermore, when I consider the macro-to-yields interaction represented by the upper right 3×3 block of $\boldsymbol{F}^*$, GDP growth has a significant effect on the slope factor while the Federal funds rate significantly influences the yields through both the level and slope factor. Estimates of the regime-switching mean are similar to those of the MS-DNS model. The mean of the slope factor in the second state is substantially lower than in the first state, indicating a steeper curve in the second regime. Lastly, both transition probabilities are significant and both states are very persistent as well.

Table 9: Estimates of the MS-DRA model

	Level_{t-1}	Slope_{t-1}	Curvature_{t-1}	GDP_{t-1}	FFR_{t-1}	CPI_{t-1}	$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_2$	λ	p_{11}	p_{22}
Level_t $(\beta_{1,t})$	**0.935**			-0.029	**0.048**	-0.016	0.328		**0.040**	**0.973**	**0.981**
	(0.052)			(0.037)	(0.041)	(0.033)	(0.022)		(0.005)	(0.009)	(0.014)
Slope_t $(\beta_{2,t})$				**0.043**	**0.335**	0.043	**-2.904**	**-5.499**			
				(0.024)	(0.086)	(0.032)	(0.094)	(0.112)			
Curvature_t $(\beta_{3,t})$			**0.907**	0.053	0.022	-0.013	-0.337				
			(0.038)	(0.021)	(0.089)	(0.052)	(0.041)				
GDP_t				**0.997**			**1.434**				
				(0.017)			(0.036)				
FFR_t					**1.000**		**0.868**				
					(0.054)		(0.044)				
CPI_t						**0.982**	**0.257**				
						(0.014)	(0.048)				

Note: This table reports a selection of the parameter estimates of the MS-DRA model. The table presents estimates of the autoregressive coefficient matrix $\boldsymbol{F}^*$, vector of means during periods of normal interest rates ($\boldsymbol{\mu}_1$) and low interest rates ($\boldsymbol{\mu}_2$), decay parameter λ and transition probabilities p_{11} and p_{22}. Bold entries denote parameter estimates significant at the five percent level. Standard errors appear in parentheses.

I evaluate the in-sample fit of the MS-DRA model by examining the diagnostics of the Rosenblatt residuals, which are shown in the fourth column of Table 5 in Section 4.1. The MS-DRA model performs substantially better than the DRA model as normality is only rejected for 2 out the 8 maturities. The in-sample gain is most profound at the short and middle end of the yield curve (except for the 36-month yield). Relative to the MS-DNS model, including macro-economic indicators as state variables improves the in-sample fit for 5 out of the 8 maturities. The MS-DRA model particularly fits the short end of the curve better. In terms of serial correlation,

[21] The own-lag coefficient of FFR_t rounds to 1.00 but is actually slightly less than one (0.9998).

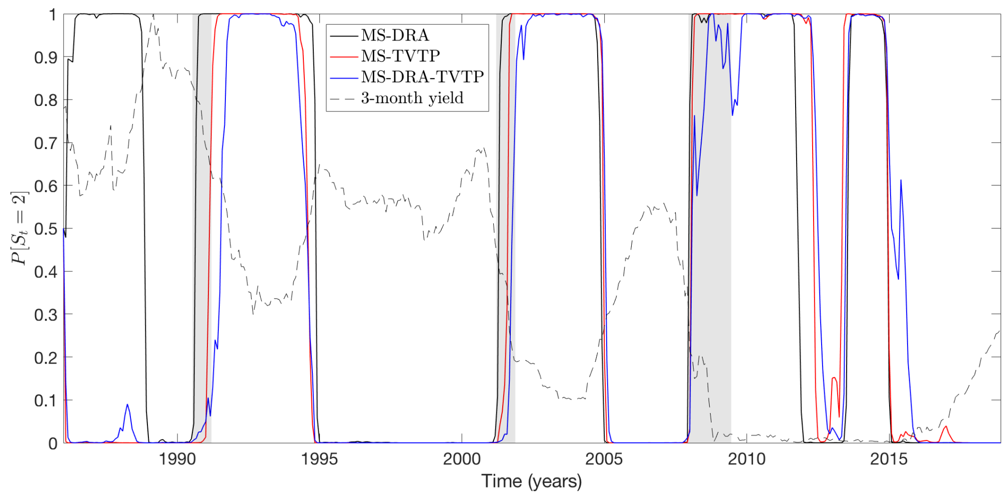
the Ljung-Box statistics of the MS-DRA model are higher for all maturities suggesting that the Rosenblatt residuals are indeed i.i.d.

Table 7 of Section 4.2 shows the performance of the MS-DRA model in terms of log-likelihood, AIC and BIC values, and LR test statistics. The improvement of the MS-DRA model over the DRA model is substantial and of a similar magnitude as that of the MS-DNS model over the DNS model. Relative to the MS-DNS model, the log-likelihood value of the MS-DRA model slightly improves. However, compared to the increase in log-likelihood value when one includes regime-switching in the DRA model, the improvement of the MS-DRA model over the MS-DNS model is not much. The same result can be found from the AIC and BIC values. The MS-DRA model, however, still significantly improves the MS-DNS model indicated by the high LR test statistic of 272.7. This suggests that the inclusion of macro-economic indicators as state variables does significantly improve the in-sample performance of the MS-DNS model.

In Figure 8, I show filtered probabilities of the second regime obtained from the MS-DRA model. The second regime is still very persistent as regime-switches occur infrequently. The filtered state probabilities of the MS-DRA model are, for a large extent, similar to those of the MS-DNS model. The former model recognizes the start of all three major crises during the sample period. The extreme low interest rate period after 2010 is also captured by the MS-DRA model. The effect of the macro-economic indicators is most profound during the beginning of the sample period until 1989. During this period the Federal funds rate, which significantly influences the slope factor in the MS-DRA model, was quite volatile and mostly decreasing. Hence, the MS-DRA model assigns a considerably higher state probability for the second regime than the MS-DNS model during the same period.

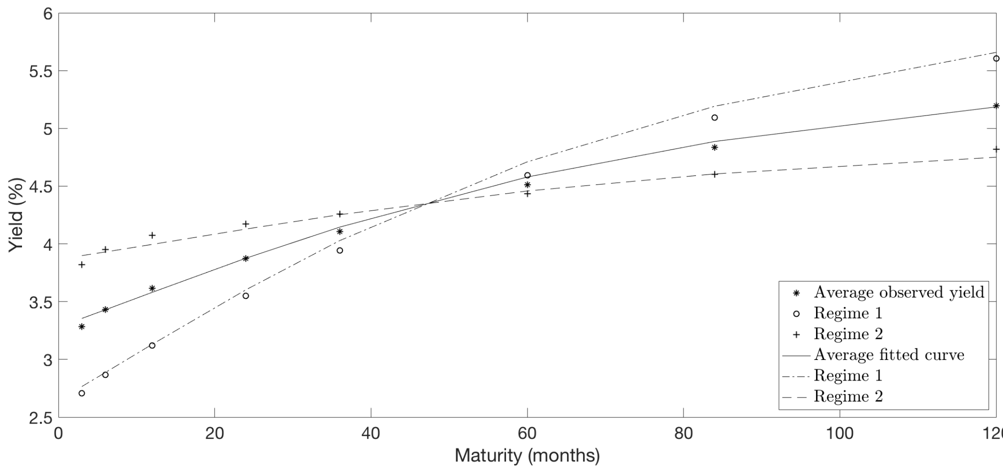
Lastly, I examine the in-sample performance of the MS-DRA model within each regime in Figure 9. The regime-dependent curves are obtained in a similar way as for the MS-DNS model in Figure 6 of Section 4.3. The yield curve is, again, substantially steeper in the low interest rate regime than in the normal regime. However, relative to the MS-DNS model the average fitted short rate in the normal regime is 0.79% lower, whereas the average fitted short rate in the low interest rate regime is 1.03% higher. This is most likely because the MS-DRA model recognizes the period before 1990 as a low interest rate regime, while interest rates were actually quite high during that time compared to interest rate levels later in the sample period.

Figure 8: Filtered probabilities second regime macro regime-switching models



Note: This figure shows filtered probabilities for the second regime, a regime with low interest rates, based on the three macro regime-switching models. It also plots the 3-month yield (dashed line), normalized between 0 and 1. Shaded areas correspond to NBER recession periods.

Figure 9: Average fit of the MS-DRA model



Note: This figure shows the average fit of the MS-DRA model. I show the average fitted yield curve over the full sample (solid line), and in each regime (dashed line). Observed yields are presented as dots, stars and crosses.

4.5 MS-TVTP: Time-Varying Transition Probabilities

I relax the assumption of a constant transition probability matrix and allow the transition probabilities to vary over time, as shown in equation (2.15) of Section 2.2.2. The parameter estimates of the autoregressive matrix $\boldsymbol{F}$ and vector of means in both regimes are similar to those of the MS-DNS model. Hence, I only present estimates of the logit model which are used to compute the TVTPs. One caveat in interpreting the coefficients in the MS-TVTP (and MS-DRA-TVTP) model is the nonlinearity of the logit model. This causes parameter interpretation to be less straightforward. To solve this problem, I follow Kole and Van Dijk (2017) who suggest to consider the marginal effect of the change in one indicator, evaluated at specific values for all indicators. For a more elaborate discussion on how the marginal effects in the logit model are computed, I refer to Appendix B.

The first two columns of Table 10 show estimates and marginal effects of the logit specification in the MS-TVTP model. The parameter estimates reveal that an increase (decrease) in GDP growth results in a significantly higher probability of staying in the normal (low interest rate) regime in the next state. This is as expected since an increase in GDP levels indicates an expansion of the economy such that interest rate cuts are not necessary. Similarly when GDP growth shrinks, it is more likely that central banks lower interest rate levels to stimulate the economy. The Federal funds rate quite surprisingly does not influence the transition probabilities significantly. This is in contrast to the estimation results of the MS-DRA model in Table 9 of Section 4.4, where the Federal funds rate does affect both the level and slope factor. Lastly, inflation is only significant in a regime with low interest rates which correctly suggests that it is an important indicator to determine whether interest rates will remain low in the next state.

The marginal effect of a one-standard deviation change is calculated using the reference probability $\hat{p}_{jj}$ and the coefficients of the indicators. The former term is calculated based on equation (B.2) of Appendix B. Consequentially, in a two-state logit model the marginal effect of indicator k in regime j with coefficient $\gamma_{j,k}$ is calculated as $\hat{p}_{jj}(1-\hat{p}_{jj})\gamma_{j,k}$. Table 10 shows that, in general, marginal effects are quite small, ranging from -0.045 to 0.144. However, the addition of the macro-economic indicators can still substantially increase the probability of remaining in the low interest rate regime. For instance, the marginal effect of a one-standard deviation increase in CPI growth almost triples the probability of remaining in the second regime from 0.078 to 0.222.

I compare the model fit of the MS-TVTP model relative to the MS-DNS and MS-DRA model in the fifth column of Table 5 in Section 4.1. Allowing for TVTPs in the MS-DNS model improves the performance for 5 out of the 8 maturities. In addition, the MS-TVTP model

Table 10: Estimates and marginal effects for the TVTPs - logit model

	MS-TVTP		MS-DRA-TVTP	
	$S_t = 1$	$S_t = 2$	$S_t = 1$	$S_t = 2$
Constant	**1.010**	**1.529**	**6.223**	**2.101**
	[0.044]	[0.110]	[0.086]	[0.051]
GDP Growth	**1.834**	**−0.626**	**1.147**	**−0.028**
	[0.081]	[−0.045]	[0.016]	[−0.001]
FFR	0.420	0.368	0.323	0.493
	[0.018]	[0.027]	[0.005]	[0.012]
CPI Growth	−0.292	**2.000**	−0.327	**1.141**
	[−0.013]	[0.144]	[−0.005]	[0.028]
$\hat{p}_{jj}$	0.954	0.078	0.986	0.025
$\bar{p}_{jj}$	0.928	0.953	0.922	0.934

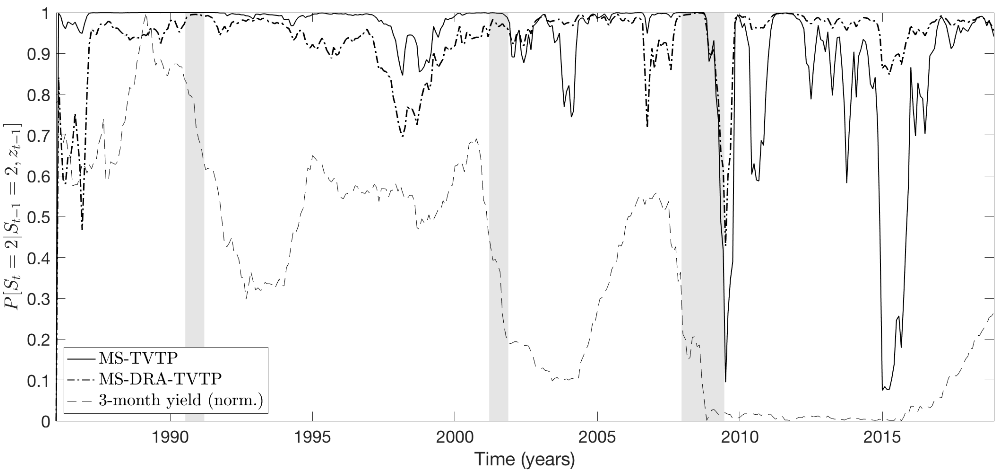
Note: This table reports estimates and marginal effects of the logit model which are used to compute the time-varying transition probabilities in the MS-TVTP and MS-DRA-TVTP model. It also presents the average transition probability $\bar{p}_{jj}$ in regime $j \in \{1,2\}$. The marginal effects, presented in brackets, are calculated using the average forecast probability $\hat{p}_{jj}$, shown in equation (B.2) of Appendix B. The marginal effect of indicator k in regime j is obtained as $\hat{p}_{jj}(1\text{-}\hat{p}_{jj})\gamma_{j,k}$. Bold entries denote parameter estimates significant at the five percent level.

performs slightly better than the MS-DRA model. The Jarque-Bera test indicates a higher p-value for the former model for 6 out of the 8 maturities. A similar pattern is found in terms of Ljung-Box statistics where the MS-TVTP model outperforms both the MS-DNS and MS-DRA model for all maturities. However the log-likelihood, AIC and BIC values in Table 7 of Section 4.2 reveal that the improvement of the MS-TVTP model is lower compared to the MS-DRA model. This is also confirmed by the significant but lower LR statistic of 163.30.

The filtered probability for the low interest rate regime obtained from the MS-TVTP model is shown in Figure 8 of Section 4.4. It is evident that the filtered probability is similar to that of the MS-DRA model. The second regime is very persistent and the MS-TVTP model is able to identify the start of the three recession periods as well as the beginning of the extreme low interest rate period in 2008.

I analyze the performance of the MS-TVTP model in more detail by examining the TVTP p_{22} which is shown in Figure 10. The MS-TVTP model is correctly able to recognize the historically low interest rate levels since 2008 but is not able to model the persistence of the low interest rates during this period. For instance, the probability of remaining in the low interest rate regime is, incorrectly, very low according to the MS-TVTP model given the first large spike. A possible explanation for this finding is that the MS-TVTP model expects interest rates to exit the ZLB in the near future since interest rates have not been decreasing at such a quick pace during the sample period. However, the second large spike around 2015 perfectly projects the development of the short rate. Shortly after p_{22} is close to zero, the short rate increases substantially from around 0% to 0.27% and, hence, exits the low interest rate regime. Therefore, it is credible to allow the transition probabilities to vary over time. Lastly, in unreported results I find the average in-sample fit of the MS-TVTP model in each regime to be similar to that of the MS-DNS model, as shown in Figure 6 of Section 4.3.

Figure 10: Transition probability p_{22} of the MS-TVTP and MS-DRA-TVTP model



Note: This figure shows the time-varying transition probability p_{22} of the two macro regime-switching models. It also plots the 3-month yield (dashed line), normalized between 0 and 1. Shaded areas correspond to NBER recession periods.

4.6 MS-DRA-TVTP: Macro-Economic State Variables and Time-Varying Transition Probabilities

Given the promising results of both the MS-DRA and MS-TVTP model, I examine in more detail the performance of the MS-DRA-TVTP model which introduces macro-economic indicators as state variables and allows the transition probabilities to vary over time. I find that the parameter estimates in the transition equation are similar to those of both the MS-DRA and MS-TVTP model.[22] Therefore, I only present estimates and marginal effects of the logit model.

The third and fourth column of Table 10 in Section 4.5 show that the parameter estimates of the TVTPs are similar to those of the MS-TVTP model: GDP growth is a significant predictor of the next state in both regimes, and inflation is only significant in the low interest rate regime. The magnitude of the marginal effects is slightly smaller in the MS-DRA-TVTP model, ranging from -0.005 to 0.086. Although the size of these effects is minor, the marginal effect of a one-standard deviation increase in CPI growth still doubles the probability of a low interest rate regime occurring in the next state from 0.025 to 0.053. The parameter estimates of the macro-economic indicators are quite encouraging. In the MS-DRA-TVTP model, the Federal funds rate significantly affects the yields within each regime through the transition equation, whereas GDP

[22]The only difference is that the macro-to-yields interaction is less apparent in the MS-DRA-TVTP model compared to the MS-DRA model. This is because the Federal funds rate is the only significant indicator in the former model.

growth and inflation significantly influence the yields across each regime through the transition probabilities.

The Rosenblatt residuals of the MS-DRA-TVTP model, which are shown in the last column of Table 5 in Section 4.1, show that the in-sample improvement is, however, only marginal relative to either the MS-DRA or MS-TVTP model. Compared to the MS-DRA and MS-TVTP model, the p-values for the Jarque-Bera test of the combined model are smaller for 1 and 3 out of the 8 maturities, respectively. This finding is confirmed by the minor increase in log-likelihood value and slight decrease in the AIC and BIC value of the MS-DRA-TVTP model, shown in Table 7 of Section 4.2.

The filtered probability of the MS-DRA-TVTP model for the low interest rate regime is shown in Figure 8 of Section 4.4. The state probabilities of the combined model are similar to those of the previous two macro regime-switching models. Hence, it appears that adding macro-economic indicators as state variables as well as through the TVTPs does not further capture the behavior of the short rate during low interest rate periods.

The TVTP of the MS-DRA-TVTP model of remaining in a regime with low interest rates, p_{22}, is shown in Figure 10 of Section 4.5. The TVTP of the combined model is far less extreme than that of the MS-TVTP model. However, the former model is also not able to capture the persistence of the low interest rates during the end of 2008. The probability of staying in the low interest rate regime at the end of 2008 is, correctly, higher for the MS-DRA-TVTP model. However, it also gives a considerable higher probability of staying in the low interest rate regime at the end of 2015. Lastly, in unreported results I find the average in-sample fit of the MS-DRA-TVTP model in each regime to be similar to that of the MS-DNS and MS-TVTP model, as shown in Figure 6 of Section 4.3.

5 Out-of-sample Forecasting

A proper term structure model should not only be able to fit the yield curve well in-sample but also generate reliable predictions of future interest rates out-of-sample. Given the promising in-sample performance of the regime-switching models, I now study whether these models are also able to forecast well out-of-sample.

I recursively estimate all models, using data from January 1986 to December 2007, and use a rolling window to generate forecasts from January 2008 to December 2018. In this way, I examine whether the regime-switching models are able to forecast the extreme low interest rate levels starting around 2008. For the baseline DNS model, forecasting is relatively straightforward as future yields only depend on the latent Nelson-Siegel factors. Hence, it is sufficient to compute the h-month ahead forecast of $\hat{\boldsymbol{\beta}}_T$ which is given by

$$\hat{\boldsymbol{\beta}}_{T+h} = [I - \hat{\boldsymbol{F}}^h][I - \hat{\boldsymbol{F}}]^{-1}\hat{\boldsymbol{\mu}} + \hat{\boldsymbol{F}}^h\hat{\boldsymbol{\beta}}_T. \tag{5.1}$$

The level and curvature factor in the MS-DNS model are computed similarly as in (5.1). However, $\hat{\boldsymbol{F}}$ and $\hat{\boldsymbol{\mu}}$ are now set to the autoregressive coefficient and mean of the respective factor, and I is a scalar equal to 1. One obtains the h-month ahead forecast for the regime-switching slope factor by taking its expectation at time $T + h$:

$$\begin{aligned}\hat{\beta}_{2,T+h} &= E[\beta_{2,T+h}|I_T] \\ &= E[\mu_{2,S_{T+h}} + \eta_{2,T+h}|I_T] \\ &= P[S_{T+h} = 1|I_T] \cdot \hat{\mu}_{2,1} + P[S_{T+h} = 2|I_T] \cdot \hat{\mu}_{2,2}.\end{aligned} \tag{5.2}$$

Afterwards, Hamilton (1995) proposes to construct forecasts of the state probabilities at time $T + h$ as

$$\hat{\pi}_{T+h|T} = \hat{P}^h\hat{\pi}_{T|T},$$

where $\hat{\pi}_{T|T}$ is found from equation (2.29). In the model extensions with macro-economic indicators, the dimension of $\hat{\boldsymbol{\beta}}_T$ is extended appropriately to include the indicators as well. Consequently for the DRA, MS-DRA and MS-DRA-TVTP model, I use $\hat{\boldsymbol{F}}^*$ as given in equation (2.16) of Section 2.2.2 to construct $\hat{\boldsymbol{\beta}}_{T+h}$. Furthermore in the MS-TVTP and MS-DRA-TVTP model, the TVTP matrix $\hat{P}_T$ as given in equation (2.12) of Section 2.2.2 is used to obtain $\hat{\pi}_{T+h|T}$. The h-month ahead forecast of the latent factors is then used to construct h-month ahead yield predictions as

$$\hat{\boldsymbol{y}}_{T+h} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{T+h}.$$

I compare the forecasts of the DNS model and its regime-switching extensions with those of a random walk (RW). Duffee (2002) argues that most affine term structure models already have trouble beating this naive model which does not use any past information to construct its forecasts. The RW is given by

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \varepsilon_t(\tau_i), \qquad \varepsilon_t(\tau_i) \sim N(0, \sigma^2(\tau_i)),$$

where $y_t(\tau_i)$ is the yield at time t with maturity τ_i, with $i = 1, .., N$ the number of different maturities. The h-month ahead forecast is then simply obtained by taking the yield of the last in-sample period. That is,

$$\hat{y}_{T+h}(\tau_i) = y_T(\tau_i).$$

I also formally test whether the baseline DNS model and its regime-switching extensions are able to produce significantly better forecasts than the RW by performing the test of Diebold and Mariano (2002) (DM) of equal forecast accuracy. In calculating the loss differentials, I assume that the relevant loss function is the squared forecast error and the DM-statistic is standard normally distributed. Diebold and Mariano (2002) note that the DM-test is not appropriate when one compares the forecast accuracy of nested models. However, recall that the MS-DNS model is not fully nested within the DNS model due to the imposed restriction that f_{22} is equal to zero.[23]

Table 11 shows root mean squared errors (RMSEs) of the RW, and ratios of the RMSEs of the Nelson-Siegel models relative to the RW, for forecast horizons h = 1, 6 and 12 months. A ratio smaller than 1 indicates that the Nelson-Siegel model forecasts the term structure better than the RW. At the 1-month horizon, the baseline DNS model already produces considerable better forecasts at the short end of the yield curve than the RW. This is consistent with Diebold and Li (2006) who also find that the DNS model with AR(1) factor dynamics is able to forecast the short end of the curve better. The inclusion of regime-switching in the DNS model only improves the prediction accuracy of the 3-month and 36-month yield. According to the DM-statistics which are shown in Table C.14 of the Appendix, the MS-DNS model predicts the long end of the yield curve significantly worse compared to the RW. When one includes macro-economic indicators, the improvement of the MS-DNS model is also negligible at the short end,

[23] In case of the MS-DRA and MS-TVTP model, which are nested within the MS-DNS model, Giacomini and White (2006) prove that the DM-test is valid when one uses a rolling window to construct forecasts, as is done in this thesis.

whereas the macro regime-switching models perform significantly worse at the long end of the curve.

When the forecast horizon lengthens, the performance of the (macro) regime-switching models improves noticeably. At the 6-month horizon, the MS-DNS model produces better forecasts for 6 out of the 8 maturities compared to the DNS model, and it outperforms the RW for 4 out of the 8 maturities. The effect of macro-economic indicators is also more profound at longer horizons. The MS-DRA and MS-TVTP model outperform the MS-DNS model for 4 and 6 out of the 8 maturities, and the RW for 3 and 4 out of 8 the maturities, respectively. The difference in RMSE between the (macro) regime-switching and DNS models at the longer end of the curve is also considerably smaller at longer horizons.

Similarly, the 12-month ahead forecasts of the (macro) regime-switching models outperform those of the baseline DNS model at all maturities. For completeness, Table C.15 and Table C.16 of the Appendix present DM-statistics of the extended Nelson-Siegel models against the DNS model and of the regime-switching models against each other, respectively. Table C.15 shows that the forecasting performance of all the regime-switching models improves especially for the 3-month yield. Furthermore, from Table C.16 it is evident that the MS-DRA-TVTP model does not further improve the forecasting performance compared to either the MS-DRA or MS-TVTP model, at all horizons. In general, the MS-TVTP model forecasts the best out of all the regime-switching models: although it does not beat the baseline DNS model at the 1-month horizon, its performance is much more promising at longer horizons.

To better understand the predictive performance of the regime-switching models over time, I plot the cumulative squared forecast errors, averaged over all maturities, in Figure 11. Consistent with the forecasting results in Table 11, the RW performs the best at a forecast horizon of 1 month. At longer horizons, the regime-switching models start to outperform both the RW and the DNS model. From the evolution of the squared forecast errors, it is evident that the regime-switching models are better able to forecast the beginning of the out-of-sample period, when short-term interest rates are reaching the ZLB. Finally, there is a sudden increase in the squared forecast error of both the MS-DRA (at the end of 2013) and MS-DRA-TVTP model (at the end of 2014). This is most likely because the filtered probability for the low interest rate regime is, incorrectly, close to zero for both models during the period 2013 - 2014, as can be seen from Figure 8 of Section 4.4.

Since the regime-switching models forecast the short end of the yield curve particularly well, I also show the cumulative squared forecast errors, averaged over the short end of the curve, in Figure C.14 of the Appendix. For the short end, I take the 3-month, 6-month and

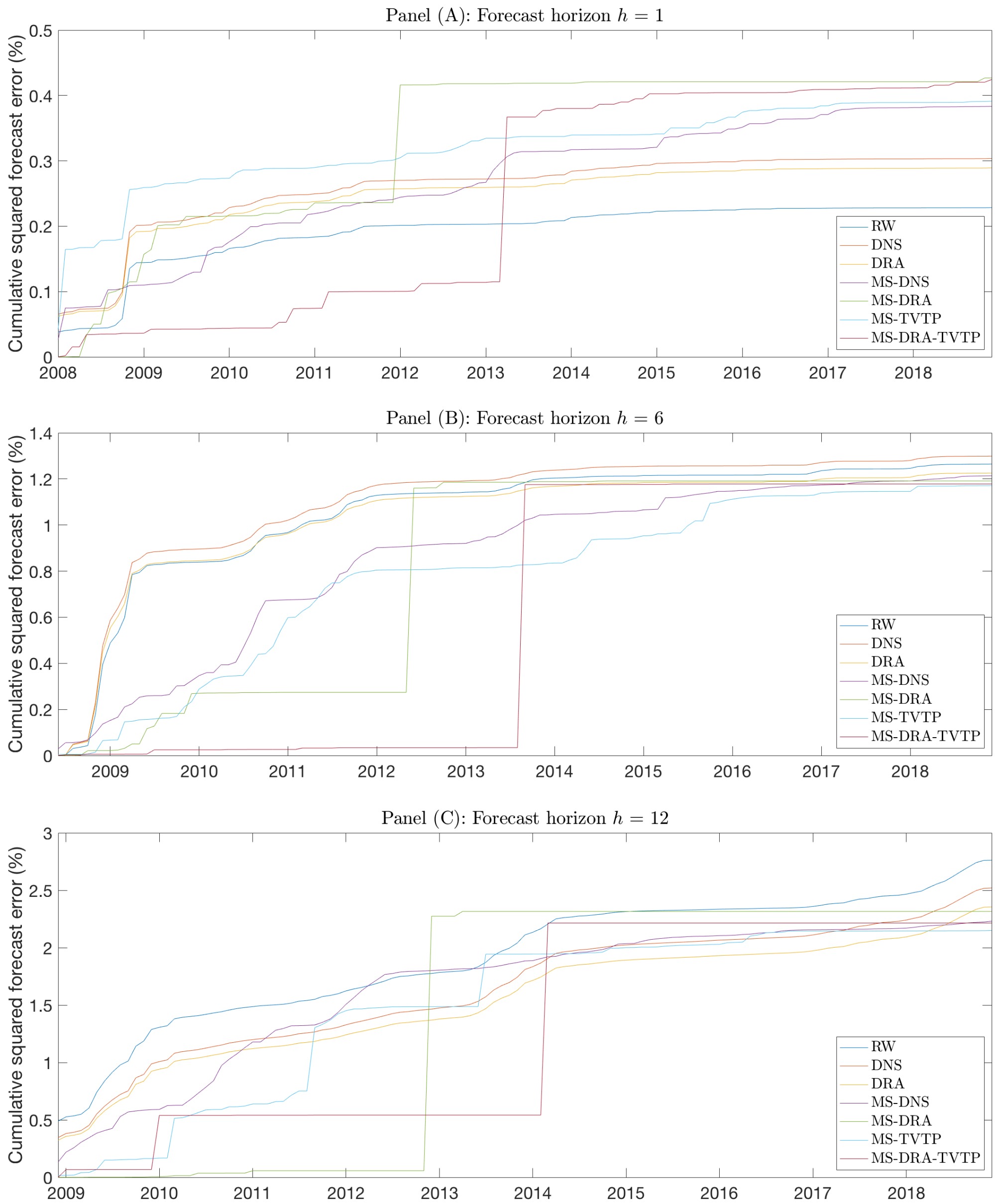
12-month yield. The outperformance of the regime-switching models relative to the RW is much more evident for short-term yields. Similar to the squared forecast errors over all maturities, the regime-switching models are especially able to better forecast the start of the extreme low interest rate period at the end of 2008.

Table 11: Root mean squared errors

Maturity (months)	RW (in %)	DNS	DRA	MS-DNS	MS-DRA	MS-TVTP	MS-DRA-TVTP
Panel (A): 1-month horizon							
3	0.53	**0.94**	**0.94**	**0.91**	**0.92**	**0.90**	**0.93**
6	0.51	**0.84**	**0.82**	**0.96**	**0.93**	**0.91**	**0.94**
12	0.47	**0.96**	**0.95**	**0.97**	**0.98**	**0.96**	**0.97**
24	0.41	**0.90**	**0.88**	1.05	1.05	1.01	1.08
36	0.37	1.45	1.47	1.32	1.30	1.27	1.32
60	0.32	1.57*	1.53*	1.75*	1.75*	1.63	1.68*
84	0.35	1.41	1.38	1.83*	1.83*	1.69*	1.70*
120	0.35	1.60*	1.62*	2.00*	2.56*	2.43*	2.55*
Panel (B): 6-month horizon							
3	1.21	**0.95**	**0.90***	**0.88***	**0.83***	**0.80***	**0.84***
6	1.18	**0.86***	**0.82***	**0.84***	**0.80***	**0.77***	**0.80***
12	1.12	**0.81***	**0.77***	**0.86***	**0.85***	**0.83***	**0.85***
24	0.99	1.07	1.11	**0.97**	1.02	**0.93**	**0.96**
36	0.91	1.04	1.03	1.03	1.04	1.02	1.03
60	0.83	1.21	1.22	1.05	1.05	1.05	1.06
84	0.83	1.18	1.17	1.23	1.22	1.21	1.22
120	0.82	1.25	1.27	1.23	1.24	1.39	1.24
Panel (C): 12-month horizon							
3	1.47	**0.91***	**0.87***	**0.81***	**0.78***	**0.71***	**0.76***
6	1.40	**0.90***	**0.84***	**0.88***	**0.87***	**0.80***	**0.82***
12	1.30	**0.92***	**0.87***	**0.90***	**0.88***	**0.82***	**0.82***
24	1.21	1.08	1.05	1.05	1.04	1.01	1.04
36	1.23	1.16	1.12	1.16	1.15	1.13	1.14
60	1.30	1.12	1.13	1.09	1.13	1.12	1.13
84	1.33	1.17	1.15	1.07	1.16	1.13	1.14
120	1.29	1.25	1.28	1.13	1.24	1.20	1.22

Note: This table reports the root mean squared error (RMSE) of the 1-month (Panel (A)), 6-month (Panel (B)) and 12-month (Panel (C)) ahead forecasts of the random walk (*RW*) in percentages, for all maturities. The other columns present RMSEs of the Nelson-Siegel models *relative* to the random walk. A value smaller than 1 is highlighted in bold and indicates an outperformance of the model relative to the random walk. An asterisk (*) denotes significant outperformance (underperformance) at the five percent level relative to the random walk when the RMSE is lower (higher) than 1, based on the test of equal forecast accuracy of Diebold and Mariano (2002) shown in Table C.14 of the Appendix. In calculating the loss differentials of the Diebold and Mariano (2002) test statistic, I assume that the relevant loss function is the squared forecast error and that the test statistic is standard normally distributed.

Figure 11: Cumulative squared forecast error - all maturities



Note: This figure presents the cumulative squared forecast error, averaged over all maturities, for the random walk (RW) and the Nelson-Siegel models, measured in percentages. Panel (A), (B) and (C) show the cumulative squared forecast error for a forecast horizon h = 1, 6 and 12, respectively.

6 Conclusion

This thesis introduces regime-switching in the mean of the Nelson-Siegel slope factor. I use the model of Bernadell et al. (2005) which distinguishes between a state where interest rate levels are normal and the yield curve is nearly flat, and a state where the curve is steep and short-term interest rates are near the ZLB. I find that the MS-DNS model is able to identify both regimes. The mean of the slope factor is close to zero in the normal regime and profoundly negative in the low interest rate regime. The regime-switching model fits the term structure considerably better in-sample than the baseline DNS model of Diebold and Li (2006), except for the 3-month yield. In addition, the filtered probabilities for the low interest rate regime coincide quite well with the start of recessions and with periods where the short-term interest rate is relatively low. The regime-switching model is also able to recognize the period after 2008 as a low interest rate regime, when interest rates are indeed near the ZLB. In an out-of-sample study, I find that the MS-DNS model produces superior forecasts relative to the DNS model. However, the regime-switching model is only able to forecast the short end of the yield curve better than a random walk.

I extend the MS-DNS model by linking the shape of the yield curve to the macro-economy in three ways: the macro-economic indicators enter the regime-switching model as state variables; the transition probabilities depend on the indicators and, hence, vary over time; and a combination of both where the indicators enter the model as state variables and through the transition probabilities. In general, the inclusion of macro-economic indicators slightly improves the in-sample performance of the MS-DNS model. The Federal funds rate is the most significant indicator in a model with macro-economic state variables. When I allow for TVTPs, the inflation level is the most important indicator in predicting whether interest rates remain low in the next state. The forecasting performance of the macro regime-switching models is also substantially better than the performance of the MS-DNS model, especially for the 3-month and 6-month yield. Including macro-economic indicators through both the state variables and the transition probabilities, however, does not further improve the performance relative to either one of the two approaches. I find that a regime-switching model with time-varying transition probabilities outperforms the other two macro regime-switching models, in terms of in-sample fit and forecasting performance.

Lastly, I note that the MS-DNS model and its extensions with macro-economic indicators are very sensitive to the initial parameter values. Similar to studies such as DRA, I use the two-step estimates of the DNS model as starting values for the Kim filter in the (macro) regime-switching models. However, I also estimated the macro regime-switching models with the estimates of the

MS-DNS model as starting values but found the results of these models to deviate substantially from those reported in this thesis.

The findings presented in this thesis provide several directions for future research. A natural extension is to consider a regime-switching model with three states. Historically, an inverted yield curve is known to predict lower future short-term interest rates as long-term bonds are more attractive causing the yield of these bonds to fall. Hence, the inclusion of a third state where the yield curve is inverted and the mean of the slope factor is positive would be interesting.

Second, besides the Kim filter, Kim and Nelson (1999) also propose a Bayesian approach to estimate state-space models with regime-switching which is based on the Gibbs sampler. An advantage of the Bayesian approach is that it overcomes the nonlinearity in the model due to the regime-switching by simulating from the conditional posterior distribution of the parameters in the Gibbs sampler. For instance, when one samples the Nelson-Siegel factors $\boldsymbol{\beta}_t$, all other parameters are assumed to be given such that the model becomes linear. In a Bayesian setting, parameter uncertainty is also taken into account directly by sampling from the posterior distribution. For an application of the Gibbs sampler in regime-switching term structure modelling, I refer to Zhu and Rahman (2009).

Lastly, one could consider a more sophisticated approach than a logit model to incorporate time variability in the transition probability matrix. Recent attention has been given to the class of dynamic score models such as the generalized autoregressive score (GAS) model of Creal et al. (2011). In the context of regime-switching models, an advantage of the GAS model is its intuitive way of updating the transition probabilities. At each time t, a score is given based on the difference between the likelihood in each regime, which is then scaled by the total likelihood given all the non-switching parameters. Bazzi et al. (2017) include TVTPs in the Markov-switching mean-variance component model of Doornik (2013) and find that the GAS model is able to effectively describe the dynamics of the transition probabilities over time.

References

Ang, A. and Bekaert, G. (2002). Regime Switches in Interest Rates. *Journal of Business & Economic Statistics*, 20(2):163–182.

Bauer, M. D. and Rudebusch, G. D. (2016). Monetary Policy Expectations at the Zero Lower Bound. *Journal of Money, Credit and Banking*, 48(7):1439–1465.

Bazzi, M., Blasques, F., Koopman, S. J., and Lucas, A. (2017). Time-Varying Transition Probabilities for Markov Regime-Switching Models. *Journal of Time Series Analysis*, 38(3):458–478.

Bernadell, C., Coche, J., and Nyholm, K. (2005). Yield Curve Prediction for the Strategic Investor.

Black, F. (1995). Interest Rates as Options. *The Journal of Finance*, 50(5):1371–1376.

Christensen, J. H. (2013). A Regime-Switching Model of the Yield Curve at the Zero Bound. Federal Reserve Bank of San Francisco.

Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models. *Journal of Econometrics*, 164(1):4–20.

Creal, D., Koopman, S. J., and Lucas, A. (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations. *Journal of Business & Economic Statistics*, 29(4):552–563.

Dai, Q., Singleton, K. J., and Yang, W. (2007). Regime Shifts in a Dynamic Term Structure Model of US Treasury Bond Yields. *The Review of Financial Studies*, 20(5):1669–1706.

De Pooter, M. (2007). Examining the Nelson-Siegel Class of Term Structure Models: In-Sample Fit versus Out-of-Sample Forecasting Performance. *Available at SSRN 992748.*

Diebold, F. X., Lee, J.-H., and Weinbach, G. C. (1994). Regime-Switching with Time-Varying Transition Probabilities. *Business Cycles: Durations, Dynamics, and Forecasting*, 1:144–165.

Diebold, F. X. and Li, C. (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X. and Mariano, R. S. (2002). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach. *Journal of Econometrics*, 131(1-2):309–338.

Doornik, J. A. (2013). A Markov-Switching Model with Component Structure for US GNP. *Economics Letters*, 118(2):265–268.

Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *The Journal of Finance*, 57(1):405–443.

Exterkate, P., Dijk, D. V., Heij, C., and Groenen, P. J. (2013). Forecasting the Yield Curve in a Data-Rich Environment using the Factor-Augmented Nelson-Siegel Model. *Journal of Forecasting*, 32(3):193–214.

Garcia, R. (1998). Asymptotic Null Distribution of the Likelihood Ratio Test in Markov-Switching Models. *International Economic Review*, pages 763–788.

Gelman, S. and Wilfling, B. (2009). Markov-Switching in Target Stocks during Takeover Bids. *Journal of Empirical Finance*, 16(5):745–758.

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

Gorovoi, V. and Linetsky, V. (2004). Black's Model of Interest Rates as Options, Eigenfunction Expansions and Japanese Interest Rates. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 14(1):49–78.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US Treasury Yield Curve: 1961 to the Present. *Journal of Monetary Economics*, 54(8):2291–2304.

Hamilton, J. D. (1995). Time Series Analysis. *Economic Theory. II, Princeton University Press, USA*, pages 625–630.

Hansen, B. E. (1992, 1996). The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov-Switching Model of GNP. *Journal of Applied Econometrics*, 7(S1):S61–S82.

Hevia, C., Gonzalez-Rozada, M., Sola, M., and Spagnolo, F. (2015). Estimating and Forecasting the Yield Curve using a Markov-Switching Dynamic Nelson and Siegel Model. *Journal of Applied Econometrics*, 30(6):987–1009.

Kim, C.-J. and Nelson, C. R. (1999). State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications. *MIT Press Books*, 1.

Kole, E. and Van Dijk, D. (2017). How to Identify and Forecast Bull and Bear Markets? *Journal of Applied Econometrics*, 32(1):120–139.

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

Kozicki, S. and Tinsley, P. A. (2001). Term Structure Views of Monetary Policy under Alternative Models of Agent Expectations. *Journal of Economic Dynamics and Control*, 25(1-2):149–184.

Krishnamurthy, A. and Vissing-Jorgensen, A. (2011). The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy. Technical report, National Bureau of Economic Research.

Levant, J. and Ma, J. (2017). A Dynamic Nelson-Siegel Yield Curve Model with Markov-Switching. *Economic Modelling*, 67:73–87.

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

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, pages 473–489.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

Rudebusch, G. and Svensson, L. E. (1999). Policy Rules for Inflation Targeting. In *Monetary policy rules*, pages 203–262. University of Chicago Press.

Smith, D. R. (2008). Evaluating Specification Tests for Markov-Switching Time-Series Models. *Journal of Time Series Analysis*, 29(4):629–652.

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

Taylor, J. B. (1993). Discretion versus Policy Rules in Practice. In *Carnegie-Rochester conference series on public policy*, volume 39, pages 195–214. Elsevier.

Wu, J. C. and Xia, F. D. (2016). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.

Zhu, X. and Rahman, S. (2009). A Regime-Switching Macro-Finance Model of the Term Structure. Technical report, working paper, Nanyang Technological University.[428].

A The Unconditional Mean and Covariance of the State Vector

Following equation (2.8) of the prediction step in the Kalman filter, the unconditional mean of the stationary state vector $\boldsymbol{\beta}_t$ is derived as

$$\begin{aligned}\boldsymbol{\beta}_{0|0} &= \boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta}_{0|0}, \\ &= (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{\mu}.\end{aligned}$$

From equation (2.9) of the prediction step in the Kalman filter, the unconditional covariance of $\boldsymbol{\beta}_t$ is derived as

$$\begin{aligned}\boldsymbol{P}_{0|0} &= \boldsymbol{F}\boldsymbol{P}_{0|0}\boldsymbol{F}' + \boldsymbol{\Sigma}_\eta, \\ \text{vec}(\boldsymbol{P}_{0|0}) &= \text{vec}(\boldsymbol{F}\boldsymbol{P}_{0|0}\boldsymbol{F}') + \text{vec}(\boldsymbol{\Sigma}_\eta), \\ &= (\boldsymbol{F} \otimes \boldsymbol{F})\,\text{vec}(\boldsymbol{P}_{0|0}) + \text{vec}(\boldsymbol{\Sigma}_\eta), \\ &= (\boldsymbol{I} - \boldsymbol{F} \otimes \boldsymbol{F})^{-1}\,\text{vec}(\boldsymbol{\Sigma}_\eta),\end{aligned}$$

where vec$(\cdot)$ denotes the vectorization operator.

B Marginal Effects in the Logit Model

In the logit model, the time-varying transition probabilities are modeled as

$$p_{jj,t} = P[S_t = j|S_{t-1} = j, \boldsymbol{z}_{t-1}] = \frac{\exp(\boldsymbol{z}'_{t-1}\gamma_j)}{1+\exp(\boldsymbol{z}'_{t-1}\gamma_j)}, \qquad j \in \{1,2\}. \tag{B.1}$$

The marginal effect of a change in the $k-$th variable z_k, evaluated at specific values for all variables $\bar{\boldsymbol{z}}$ is then simply obtained by taking the first derivative of (B.1) with respect to z_k:

$$\frac{\partial p_{jj}(\boldsymbol{z})}{\partial z_k}\Big|_{\boldsymbol{z}=\bar{\boldsymbol{z}}} = p_{jj}(\bar{\boldsymbol{z}})(1-p_{jj}(\bar{\boldsymbol{z}}))\gamma_{j,k},$$

where $\gamma_{j,k}$ is the coefficient on z_k in regime j.

Next, to evaluate the effect of a change in z_k on the transition probabilities, I follow Kole and Van Dijk (2017) and calculate the marginal effect of a one-standard deviation change in z_k on a certain reference probability p. Kole and Van Dijk (2017) suggest to use the average forecast probability $\hat{p}_{jj}$ as the reference probability:

$$\hat{p}_{jj} = \frac{\sum_{t=1}^{T} P[S_{t+1} = j|S_t = j, \boldsymbol{z}_{t-1}]P[S_t = j|I_t]}{\sum_{t=1}^{T} P[S_t = j|I_t]}. \tag{B.2}$$

Here, each forecast probability of staying in regime j, $P[S_{t+1} = j|S_t = j, \boldsymbol{z}_{t-1}]$, is weighted by the probability of entering regime j at time t, $P[S_t = j|I_t]$. The latter probability is obtained through the Hamilton step of the Kim filter, shown in equation (2.26) of Section 2.2.3.

C Additional Tables and Figures

Table 12: Principal component analysis

Maturity in months	First P.C.	Second P.C.	Third P.C.
3	0.36	-0.42	0.53
6	0.37	-0.39	0.23
12	0.38	-0.28	-0.18
24	0.38	-0.09	-0.44
36	0.37	0.07	-0.43
60	0.34	0.30	-0.18
84	0.32	0.44	0.12
120	0.30	0.56	0.46
% explained	97.13	2.66	0.18

Note: This table presents the factor loadings of the US Treasury yields according to a principal component analysis. For each maturity, the first eight rows show how the bond yields load on the first three principal components. The last row presents the total variation in the yield data, explained by each principal component.

Table 13: Estimates of the DNS model - VAR specification

Panel (A): Vector-autoregressive matrix $\boldsymbol{F}$ and vector of means $\boldsymbol{\mu}$

	Level$_{t-1}$	Slope$_{t-1}$	Curvature$_{t-1}$	$\boldsymbol{\mu}$
Level$_t$ ($\beta_{1,t}$)	**0.977**	0.005	**0.022**	0.045
	(0.012)	(0.017)	(0.010)	(0.050)
Slope$_t$ ($\beta_{2,t}$)	-0.006	**0.965**	0.020	-0.031
	(0.013)	(0.017)	(0.011)	(0.021)
Curvature$_t$ ($\beta_{3,t}$)	0.045	0.040	**0.905**	-0.092
	(0.036)	(0.029)	(0.024)	(0.055)

Panel (B): Covariance matrix $\boldsymbol{\Sigma}_\eta$ and decay parameter λ

	Level$_t$ ($\beta_{1,t}$)	Slope$_t$ ($\beta_{2,t}$)	Curvature$_t$ ($\beta_{3,t}$)	λ
Level$_t$ ($\beta_{1,t}$)	**0.134**	-0.124	-0.138	**0.038**
	(0.009)	(0.015)	(0.032)	(0.000)
Slope$_t$ ($\beta_{2,t}$)		**0.168**	0.13	
		(0.006)	(0.042)	
Curvature$_t$ ($\beta_{3,t}$)			**0.834**	
			(0.029)	

Note: This table reports estimates of the baseline DNS model where the latent factors follow a VAR process. Panel (A) presents estimates of the vector-autoregressive coefficient matrix $\boldsymbol{F}$ and vector of means $\boldsymbol{\mu}$. Panel (B) presents estimates of the covariance matrix $\boldsymbol{\Sigma}_\eta$ and decay parameter λ. Bold entries denote parameter estimates significant at the five percent level. Standard errors appear in parentheses.

Table 14: Diebold-Mariano statistics - random walk model

Maturity (months)	DNS	DRA	MS-DNS	MS-DRA	MS-TVTP	MS-DRA-TVTP
Panel (A): 1-month horizon						
3	**1.12**	**1.12**	**1.18**	**1.13**	**1.34**	**1.07**
6	**1.37**	**1.43**	**1.03**	**1.17**	**1.35**	**1.10**
12	**1.16**	**1.10**	**1.18**	**1.12**	**1.19**	**1.05**
24	**1.61**	**1.72**	−0.38	−0.21	−0.14	−0.22
36	−1.67	−1.74	−1.68	−0.83	−0.73	−0.69
60	−2.78*	−2.54*	−2.44*	−2.28*	−1.45	−2.00*
84	−1.85	−1.68	−2.85*	−2.33*	−2.32*	−2.04*
120	−2.99*	−3.06*	−3.78*	−2.77*	−3.15*	−2.50*
Panel (B): 6-month horizon						
3	**1.90**	**2.12***	**2.62***	**2.45***	**2.85***	**2.47***
6	**2.34***	**2.53***	**2.61***	**2.49***	**2.81***	**2.42***
12	**2.71***	**2.90***	**2.53***	**2.36***	**2.63***	**2.30***
24	−0.50	−0.66	**1.13**	−0.01	**1.22**	**1.07**
36	−0.80	−0.84	−0.11	−0.07	−0.07	−0.05
60	−1.16	−1.18	−0.11	−0.07	−0.10	−0.07
84	−1.87	−1.84	−0.60	−0.39	−0.61	−0.31
120	−1.08	−1.21	−0.71	−0.42	−1.25	−0.34
Panel (C): 12-month horizon						
3	**3.97***	**4.12***	**3.56***	**3.63***	**4.01***	**3.69***
6	**4.29***	**4.48***	**4.43***	**3.37***	**3.98***	**3.48***
12	**4.03***	**4.37***	**3.48***	**3.51***	**4.32***	**3.23***
24	−0.32	−0.22	−0.62	−0.17	−0.23	−0.68
36	−0.41	−0.36	−0.63	−0.21	−0.31	−0.77
60	−0.66	−0.68	−0.35	−0.18	−0.28	−0.73
84	−0.69	−0.61	−0.43	−0.24	−0.41	−0.77
120	−1.55	−1.74	−1.09	−0.39	−0.89	−0.92

Note: This table reports statistics of the Diebold-Mariano (DM) test for equal forecast accuracy between the Nelson-Siegel models and the random walk. Under the null hypothesis, the forecasts of both models have the same mean squared error. A positive DM-statistic is highlighted in bold and indicates that the model constructs better forecasts than the random walk. An asterisk (*) denotes significance relative to the asymptotic null distribution at the five percent level.

Table 15: Diebold-Mariano statistics - baseline DNS model

Maturity (months)	DRA	MS-DNS	MS-DRA	MS-TVTP	MS-DRA-TVTP
Panel (A): 1-month horizon					
3	**0.00**	**0.15**	**0.11**	**0.30**	−0.21
6	**0.14**	−0.25	−0.15	−0.01	−0.20
12	**0.12**	−0.05	−0.03	**0.13**	−0.09
24	**0.17**	−2.21*	−1.98*	−1.65	−1.95
36	−0.14	**1.92**	**1.74**	**1.87**	**1.94**
60	**0.51**	−0.88	−1.12	−1.84	−1.41
84	**0.91**	−1.29	−1.08	−1.06	−0.74
120	−0.13	−1.14	−0.45	−0.28	−0.93
Panel (B): 6-month horizon					
3	**0.62**	**0.88**	**0.79**	**1.30**	**0.94**
6	**0.58**	**0.62**	**0.56**	**0.70**	**0.53**
12	**0.57**	−0.37	−0.43	−0.33	−0.45
24	−0.41	**2.26***	**1.02**	**2.44***	**2.14***
36	−0.35	**1.14**	**1.21**	**1.21**	**1.27**
60	−0.32	**1.09**	**1.15**	**1.10**	**1.15**
84	**0.33**	−1.32	−1.51	−1.33	−1.67
120	−0.42	**1.66**	**1.89**	−0.37	**1.71**
Panel (C): 12-month horizon					
3	**2.04***	**2.40***	**2.65***	**2.98***	**2.18***
6	**2.04***	**2.33***	**2.18***	**1.93**	**1.04**
12	**2.08***	**2.36***	**2.77***	**2.07***	**1.05**
24	**1.69**	**2.44***	**1.73**	**1.62**	**0.76**
36	**1.48**	**0.00**	**1.51**	**1.16**	**0.46**
60	−1.03	**1.53**	**2.27***	**1.82**	**1.05**
84	**1.58**	**1.72**	**2.35***	**1.59**	**1.39**
120	−1.12	**1.70**	**3.25***	**2.57***	**2.27***

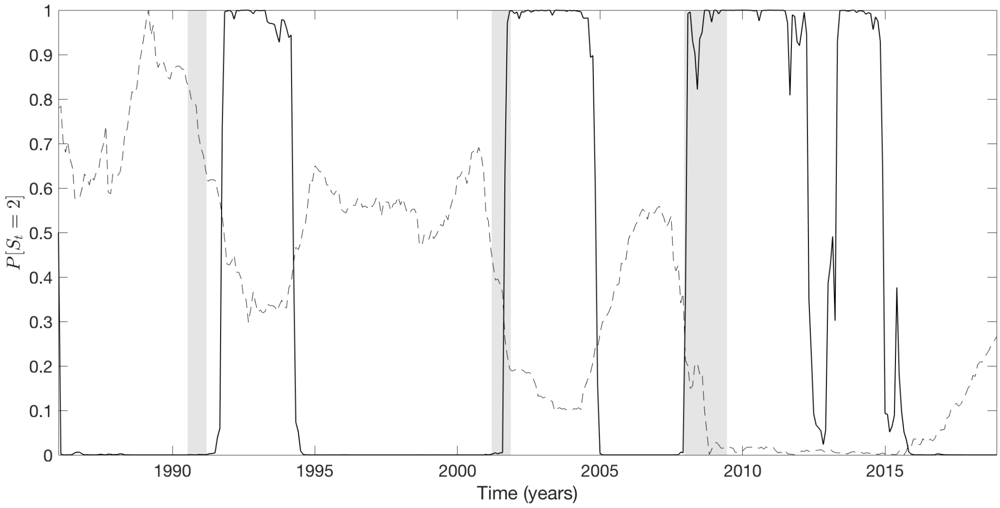
Note: This table reports statistics of the Diebold-Mariano (DM) test for equal forecast accuracy between the Nelson-Siegel models and the baseline DNS model. Under the null hypothesis, the forecasts of both models have the same mean squared error. A positive DM-statistic is highlighted in bold and indicates that the model constructs better forecasts than the DNS model. An asterisk (*) denotes significance relative to the asymptotic null distribution at the five percent level.

Table 16: Diebold-Mariano statistics - regime-switching models

	MS-DNS			MS-DRA		MS-TVTP
Maturity (months)	MS-DRA	MS-TVTP	MS-DRA-TVTP	MS-TVTP	MS-DRA-TVTP	MS-DRA-TVTP
Panel (A): 1-month horizon						
3	−0.34	**1.44**	−0.39	**1.89**	−0.35	−0.80
6	**0.44**	**0.61**	**0.37**	**1.85**	−0.36	−0.81
12	−0.35	**0.31**	−0.41	**1.66**	**1.36**	−0.88
24	**0.45**	**0.67**	**0.48**	**1.37**	−0.25	−0.97
36	**2.19***	**2.33***	**1.51**	**1.58**	−0.83	−0.95
60	**2.33***	**2.59***	**1.82**	**2.64***	**1.88**	−1.38
84	**1.82**	**1.85**	**1.32**	**0.05**	**0.88**	−0.88
120	**2.73***	**2.23***	**2.16***	−0.54	**1.90**	−1.19
Panel (B): 6-month horizon						
3	**1.09**	**2.09***	**1.06**	**2.01***	−1.47	−1.51
6	**1.05**	**2.08***	**1.09**	**2.17***	−0.84	−1.27
12	**1.09**	**2.04***	**1.15**	**1.97***	−0.83	−1.49
24	−4.01	**2.08***	**1.95**	**6.24***	**5.88***	−1.12
36	**1.64**	**1.64**	**1.85**	**0.00**	**0.71**	−1.21
60	**1.64**	**1.31**	**1.64**	−1.00	**0.00**	−1.09
84	**2.05***	−0.82	−2.52*	−0.81	**1.51**	−1.74
120	**2.59***	−1.76	−2.68*	−2.81*	**1.27**	**2.41***
Panel (C): 12-month horizon						
3	**2.02***	**2.53***	**2.32***	**2.10***	**3.29***	−1.87
6	−1.76	−0.90	**1.01**	**2.18***	**3.33***	−1.57
12	**1.61**	**2.64***	**1.22**	**2.23***	**2.21***	−1.98
24	**3.27***	**2.37***	**2.19***	**2.35***	**2.06***	−1.78
36	**3.33***	**2.49***	**1.89**	**2.48***	**2.29***	−1.87
60	**2.81***	**2.12***	**1.55**	**2.56***	**2.28***	−1.82
84	**2.96***	**1.95**	**1.63**	**2.71***	**2.13***	−1.66
120	**3.36***	**2.12***	**2.19***	**2.28***	**2.08***	−1.47

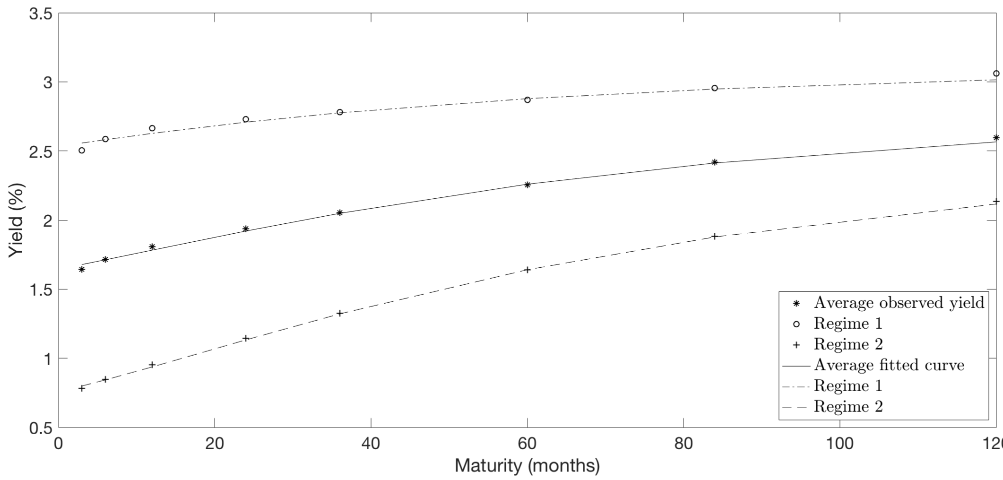
Note: This table reports statistics of the Diebold-Mariano (DM) test for equal forecast accuracy between the (macro) regime-switching models. The model in the first row is the model of which the forecast accuracy is compared against. Under the null hypothesis, the forecasts of both models have the same mean squared error. A positive DM-statistic is highlighted in bold and indicates that the model constructs better forecasts than the model shown in the first row. An asterisk (*) denotes significance relative to the asymptotic null distribution at the five percent level.

Figure 12: Filtered probabilities second regime MS-DNS model including f_{22}



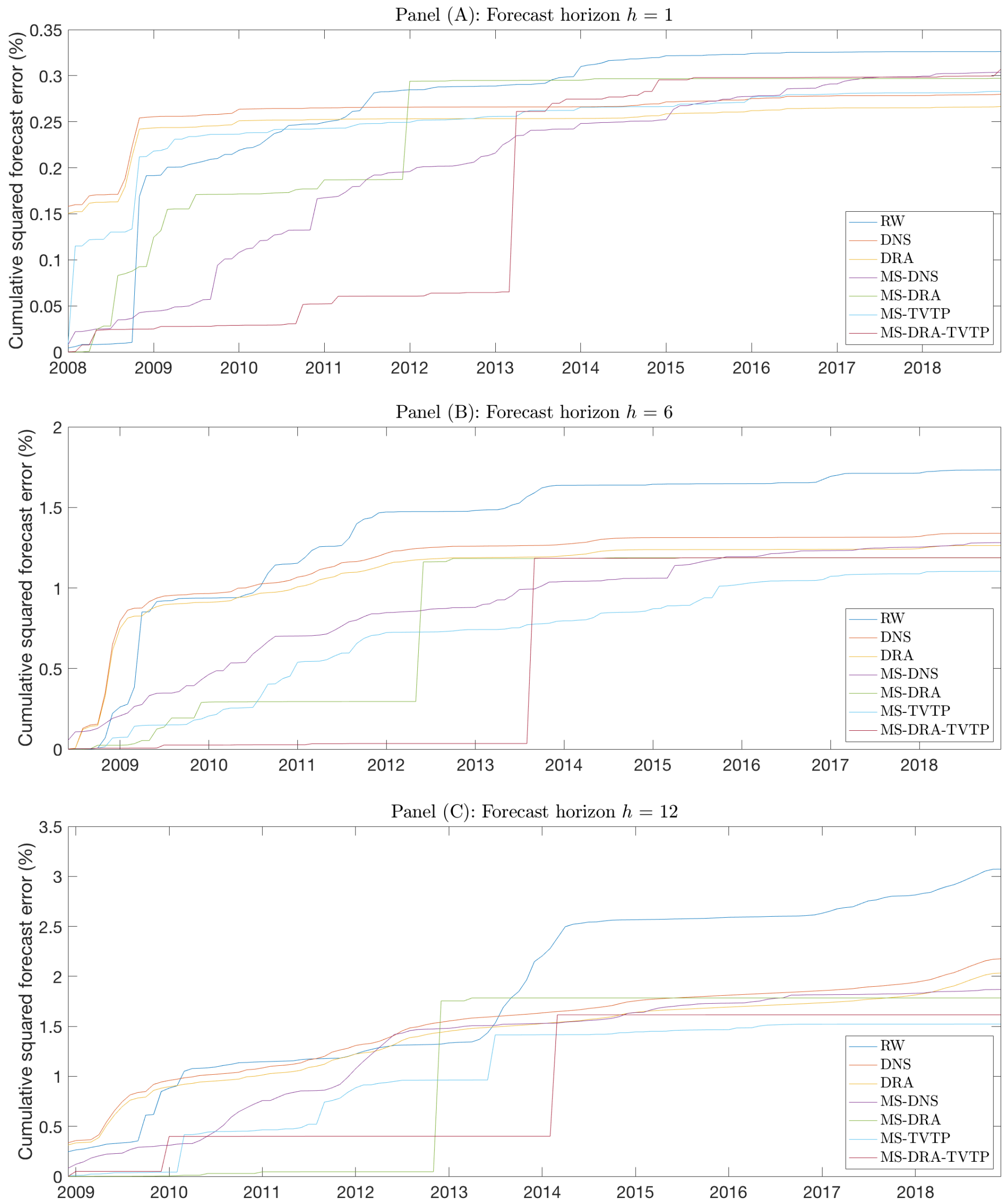
Note: This figure shows filtered probabilities for the second regime, a regime with low interest rates, based on the MS-DNS model which includes the slope factor f_{22} in the autoregressive matrix $\boldsymbol{F}$. It also plots the 3-month yield (dashed line), normalized between 0 and 1. Shaded areas correspond to NBER recession periods.

Figure 13: Average fit of the MS-DNS model - probability weighted curve



Note: This figure shows the average fit of the MS-DNS model. I show the average fitted yield curve over the full sample (solid line), and in each regime (dashed line). The regime-dependent curves are obtained by taking the probability-weighted average of the fitted yields and the filtered state probabilities in each regime. Observed yields are presented as dots, stars and crosses.

Figure 14: Cumulative squared forecast error - short end yield curve



Note: This figure presents the cumulative squared forecast error, averaged over the short end of the yield curve, for the random walk (RW) and the Nelson-Siegel models, measured in percentages. For the short end, I take the 3-month, 6-month and 12-month yield. Panel (A), (B) and (C) show the cumulative squared forecast error for a forecast horizon $h = 1$, 6 and 12, respectively.