ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Time-Varying Parameters in the Dynamic Nelson-Siegel Model

Econometrics and Management Science
Master Thesis

Erasmus Universiteit Rotterdam

Jos Buitenhuis
374725

May 22, 2017

Abstract

In this study we look into various ways to introduce time-varying parameters in the dynamic Nelson-Siegel model. We consider three extensions: (i) a time-varying loading parameter, (ii) time-varying volatility and (iii) a time-varying unconditional mean. For the introduction of a time-varying unconditional mean we consider specifications with- and without macroeconomic information. Furthermore, we combine the three extensions into a more sophisticated model. We perform an in- and out-of-sample analysis applied to zero coupon US government bond yields. We find clear evidence that the introduction of a time-varying loading parameter and time-varying volatility improves the in-sample fit, whereas evidence for the introduction of a time-varying unconditional mean is less pronounced. Furthermore, we find evidence that combining extensions improves the fit even further. In terms of predictive performance we find that time variation in the loading parameter and in the unconditional mean without the use of macroeconomic information generally improve the forecast performance on the complete yield curve, whereas the extensions of time-varying volatility only improves for the short-end of the curve. In addition we find that combining extensions leads to a model superior to all models with one specific extension and that macroeconomic information is only beneficial in periods of high volatility.

Keywords: Dynamic Nelson-Siegel, Extended Kalman Filter, Time-Varying Parameters, Macroeconomic Information

Supervisor: D.J.C. van Dijk
Co-reader: X. Gong

Contents

1 Introduction

The term structure of interest rates is the relation between interest rates/bond yields and different terms of maturities. Yields are known to be closely related over time as well as across maturities. However, short- and long-term yields usually react differently to new information shocks, which makes modelling and forecasting the term structure a challenging task. Yet, it is a very important task as it plays a crucial role in areas such as portfolio management, financial risk management, monetary policy and financial instrument pricing. It is not surprising, therefore, that a vast literature for modelling and predicting the dynamics of the yield curve has already been produced. Past work mainly focuses on either theoretically substantiated models or purely statistical models. In this study we focus on the latter type of models by considering extensions to the popular and well known Dynamic Nelson-Siegel (DNS) model.

We investigate whether the introduction of time-varying parameters in the DNS model in state space form improves in- and out-of-sample performance. We consider three extensions: (i) a time-varying loading parameter, (ii) a time-varying unconditional mean, and (iii) time-varying volatility. In addition we examine the usefulness of macroeconomic information for fitting and predicting the term structure by incorporating macroeconomic factors in the specification with a time-varying unconditional mean. Hereby we make a comparison between principal component analysis (PCA) and partial least squares (PLS) as factor construction methods. Furthermore, we examine the performance of models that combine all above mentioned extensions in a quest for finding an optimal performing model.

The literature on term structure modelling has focused for many years on affine term structure models and no-arbitrage models, which are both of theoretical nature. Early and prominent contributions on these type of models are made by Vasicek (1977), Cox, Ingersoll Jr, and Ross (1985) and Hull and White (1990). Many others proceeded on their work finding that these type of models generally do well in fitting the term structure but are not suitable for forecasting purposes. Possibly because of the latter, more recent work shifted attention back to the statistical type of models based on the work of Nelson and Siegel (1987). In this model, the yield curve is expressed in terms of three dynamically evolving exponential parameters. The Nelson-Siegel framework imposes structure on the factor loadings, making highly accurate estimation of the factors possible. Diebold and Li (2006) show that the factors can be interpreted as level, slope and curvature. Furthermore, they propose and estimate autoregressive models for the factors, and forecast the yield curve by forecasting the factors. This model is referred to as the DNS model. Their forecasting results appear to be much more accurate at longer horizons than various standard benchmark forecasts (including random walk).

Inspired by the good performance of the DNS model, many related and extending studies followed. Well known is the Svensson (1995) model where the Nelson-Siegel specification is extended by including an additional latent factor representing a second hump-shape to provide more flexibility. A similar approach is considered by De Pooter (2007), who includes an additional slope factor inspired by the four factor approximation for the forward curve of Björk and

Christensen (1999). Diebold, Rudebusch, and Aruoba (2006) put the DNS model of Diebold and Li (2006) into state space form. The state space representation follows from modelling the dynamics of the factors as a vector autoregressive process of order one and is a convenient representation for fitting and forecasting. Other work using the state space representation includes De Pooter (2007) and Yu and Zivot (2011), among many others. Associated results generally provide evidence that these models outperform multiple benchmark models in terms of forecasting accuracy.

As the state space approach of the DNS model has shown to perform well and to be easily adaptable for extensions, several studies have attempted to improve the DNS model by making it more flexible. Useful improvements are found in extensions that allow for time-varying parameters. Bianchi, Mumtaz, and Surico (2006) introduce time-varying volatility for the latent factors and find that it improves the in-sample fit. Koopman, Mallee, and Van der Wel (2010) go a step further by introducing time variation in the loading parameter and specifying the common variance as a generalized autoregressive conditional heteroscedasticity (GARCH) process, see Bollerslev (1987). They present empirical evidence of substantial improvements in within-sample goodness of fit for these extensions to the DNS model. Dijk, Koopman, Wel, and Wright (2014) extend the DNS model by allowing for a time-varying unconditional mean in the specification for the factors, although this extension is not applied in state space form. They find that this extension provides substantial gains in predictive performance. Besides the extensions with time-varying parameters, several studies extended the DNS model by incorporating macroeconomic information in an attempt to capture the interaction between yields and the macroeconomy, see Diebold et al. (2006), Mönch (2012), Yu and Zivot (2011) and Exterkate, Dijk, Heij, and Groenen (2013).

Although the literature on extensions to the DNS model is quite extensive, most studies focus on in-sample performance and fairly limited amount of effort has been put in the forecast ability of these models. This is particularly the case for the work using time-varying parameters in the DNS model in state space form. With this study we aim to fill this gap. Furthermore, the use of time-varying parameters has, to our best knowledge, never been combined with the use of macroeconomic information. In addition, we contribute to the existing literature by considering a new yield dataset and a new macroeconomic dataset. The yield dataset consists of US yield curve data for ten maturities and includes maturities up to thirty years. This is a wider set of maturities than considered in most studies and possibly results in different dynamics. The macroeconomic dataset consists of 128 macroeconomic variables and is recently composed by members of the Federal Reserve Bank of St. Louis. Moreover, we consider a time span that contains a very recent period with extremely low yields not having been studied extensively yet.

In order to reach the goals of this study we start with considering the Nelson-Siegel formulation, and modelling the factors' dynamic movements as an autoregressive process of order one. This implies almost directly a state space representation. Hereby we use constant factor loadings for the entire estimation sample. Next, we consider a model that introduces a time-varying loading parameter and a model that incorporates time-varying volatility. In terms of methodology we follow Koopman et al. (2010) for these models.

Next, we consider extensions that introduce a time-varying unconditional mean in the DNS model, using two approaches. In the first approach we take an exponential smoothing (ES) specification for the mean parameter as in Dijk et al. (2014). In the second approach we extend the model by including macroeconomic factors that are constructed from a large panel of macroeconomic variables into the state variable, following Exterkate et al. (2013). For the extraction of factors from the macro dataset we consider PCA and PLS.

Finally, we combine all types of extensions (variation in mean, loading parameter and volatility), resulting in three combined models. These models differ in terms of their approach to introduce a time-varying unconditional mean, namely (i) ES, (ii) PCA and (iii) PLS. For the estimation of the linear state space models we use Maximum Likelihood in combination with the Kalman filter, whereas we use Maximum Likelihood in combination with the extended Kalman filter for the estimation of the nonlinear state space models.

We evaluate the in-sample performance of all considered models by examining the log-likelihood, the Akaike information criterion, the Bayesian information criterion and the likelihood-ratio statistics, as well as the filtered errors. For evaluating out-of-sample performance we consider the root mean squared prediction errors (RMSPE) of every model relative to the RMSPE of the benchmark DNS model. We test significance of improvements in RMSPE using the Diebold and Mariano (1995) test.

We find clear evidence that the extensions of time-varying volatility and a time-varying loading parameter substantially increase the in-sample fit, whereas evidence for the introduction of a time-varying unconditional mean is less pronounced. We find that combining extensions leads to even better in-sample performance than the performance of individual extensions. By evaluating the predictive performance we find that the extensions with a time-varying loading parameter and a time-varying unconditional mean without the use of macroeconomic information generally result in better predictive performance for the entire yield curve, whereas the introduction of time-varying volatility only improves predictions at the short end of the curve. Moreover, we find that combining extensions leads to a model superior to all single-extension models and that the use of macroeconomic information is only beneficial in a period where volatility is high.

The remainder of this study is organized as follows. In section 2, we discuss all considered models and methodology. Section 3 describes the data and presents its characteristics. Section 4 presents the in-sample results. In section 5 we discuss the forecast procedure, the forecast evaluation and the out-of-sample results. Section 6 concludes. Section 7 discusses possibilities for further research.

2 Models and Methodology

In this section we specify the considered models and describe the associated estimation techniques. The section is split into eight subsections in order to provide a structured overview of the methodology. Section 2.1 describes the Nelson-Siegel structure and its interpretation. Section 2.2 specifies the DNS in state space form. Sections 2.3 - 2.5 describe extensions to the DNS model in state space form. The estimation of linear and nonlinear state space models is explained in sections 2.7 and 2.8, respectively.

2.1 Nelson-Siegel Factors

Nelson and Siegel (1987) introduced a factor representation of the yield curve which is nowadays still very popular. Diebold and Li (2006) modified the representation which comes down to the following.

Let $y_t^{(\tau_i)}$ be a yield series at time t for a set of maturities $\{\tau_i\}_{i=1}^N$. Then, the yield curve is described by the following formulation

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

where λ is fixed and is referred to as the loading parameter. In this formulation the yields are determined over time by the latent factors $\beta_{1,t}, \beta_{2,t}$ and $\beta_{3,t}$, while they are determined over the cross-section by the expressions presented next to the betas. These expressions are called the factor loadings. In order to obtain a better knowledge about the interpretation of this formulation it is useful to examine these expressions. The first expression takes the value 1, which suggests that $\beta_{1,t}$ can be seen as an additive (constant) factor that influences the yields at all maturities. Therefore it can be interpreted as the level factor. For the second and third expression it is useful to check the limits of the Nelson-Siegel formulation. These are given by

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

From the limits it is notable that the short end of the curve is affected by the factors $\beta_{1,t}$ and $\beta_{2,t}$, while the long end of the curve is only affected by the factor $\beta_{1,t}$. Hence, a change in $\beta_{2,t}$ has a larger effect on the short-term yields than on the long-term yields and therefore determines the slope of the curve. In fact, the slope of the yield curve can be defined as the long end of the yield curve minus the short end of the yield curve. Using the results of (2), we obtain $y_t^{(\infty)} - y_t^{(0)} = -\beta_{2,t}$. This result confirms that $\beta_{2,t}$ can be interpreted as the slope factor. Furthermore, it can be seen from (2) that the limit to zero and to infinity are both equal to 0 for the third expression in (1). Given that this expression is a concave function of τ_i, we know that it influences the medium term yields the most. Therefore, this expression determines the shape of the curve and is called the curvature factor. To clarify the interpretation of the factors figure 1 displays the factor loadings as a function of the maturity τ_i for $\lambda = 0.04$. Another

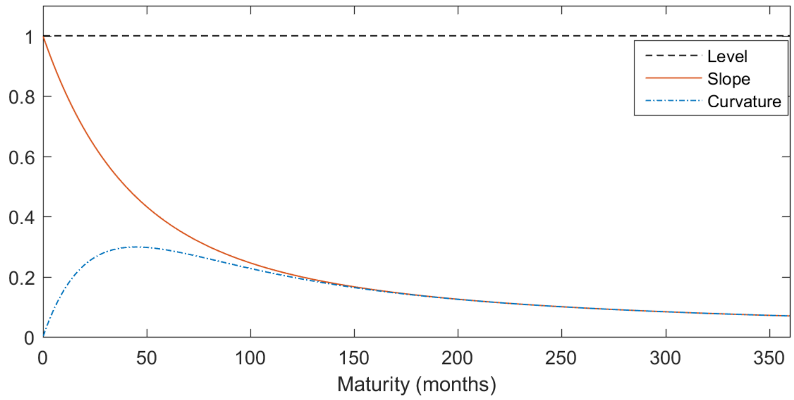


Figure 1: Nelson-Siegel factor loadings. This figure displays the loadings for the level, slope and curvature factors in the Nelson-Siegel equation as a function of maturity for $\lambda = 0.04$.

important element of the Nelson-Siegel formula in (1) that determines the shape of the curve is the loading parameter λ. This parameter determines the exponential decay in the slope loading and the maximum at which the curvature loading reaches its maximum. As discussed in section 3 is known to take a variety of shapes over time. The formulation in (1) is able to replicate these shapes very well and is therefore widely used to fit the yield curve.

2.2 DNS in State Space Form

As mentioned in section 2.1, the factors $\beta_t = \{\beta_{1,t}, \beta_{2,t}, \beta_{3,t}\}$ in the Nelson-Siegel formulation determine the yield curve over time. This means that if one is able to forecast these factors, one is directly able to forecast the yield curve as well. Diebold and Li (2006) recognize that the factors in the Nelson-Siegel equation are strongly correlated over time, which suggests that they are forecastable. Given this finding they propose a 2-step approach for forecasting the yield curve. First, they estimate β_t by applying least squares over the cross section in (1) for all periods in time $t = 1, \ldots, T$, such that they obtain a time series of T estimates for β_t. Second, they complete the model by autoregressive specifications for the dynamics of the factors. More specifically, they suggest separate univariate autoregressive processes of order one, given by

$$\beta_{j,t+1} = \mu_j + \phi_j \left(\beta_{j,t} - \mu_j\right) + \epsilon_{j,t} \quad \text{for } j = 1, 2, 3, \tag{3}$$

where μ_j is the unconditional mean of β_j and it is assumed that the disturbances $\epsilon_{j,t+1}$ are normally distributed with mean zero and variance σ_j^2, and are mutually and serially independent at all time periods. To obtain the h-period ahead yield forecast $y_{j,t+h}$, one must first obtain the h-period ahead factor forecast $\beta_{j,t+h}$ by iterating (3) h times and then substitute the constructed factor forecast into (1). Hence, this approach facilitates a simple way of forecasting yields. Furthermore, Diebold and Li (2006) show that this approach is able to produce accurate forecasts.

Diebold et al. (2006) recognize that the Nelson-Siegel model can be framed into a state space model, where the factors are treated as latent factors. A state space model is a way of representing a time series model in terms of a system of equations that describes how the observed time series is linked to latent factors, and how the latent factors evolve over time. The general state space model used in this study is presented as follows

$$y_t = Hx_t + \omega_t, \qquad \omega_t \sim \mathcal{N}(0, R), \tag{4}$$

$$x_{t+1} = C + Ax_t + \nu_t, \qquad \nu_t \sim \mathcal{N}(0, Q), \tag{5}$$

for $t = 1, \ldots, T$, where y_t are the observed time series of interest, x_t is the state vector, H, C and A are coefficient matrices and R and Q are variance matrices. Equation (4) is called the measurement equation or observation equation and describes the relation between the observed time series and the state vector. Equation (5) is called the state equation and describes how the latent factors evolve over time.

Moving back to the specification of the DNS model from Diebold and Li (2006), it is clearly visible that this specification has the same structure as a state space model. Following Diebold et al. (2006), but letting the factors follow a separate autoregressive process of order one, the state equation is then given by

$$\begin{bmatrix} \beta_{1,t+1} - \mu_1 \\ \beta_{2,t+1} - \mu_2 \\ \beta_{3,t+1} - \mu_3 \end{bmatrix} = \begin{bmatrix} \phi_{11} & 0 & 0 \\ 0 & \phi_{22} & 0 \\ 0 & 0 & \phi_{33} \end{bmatrix} \begin{bmatrix} \beta_{1,t} - \mu_1 \\ \beta_{2,t} - \mu_2 \\ \beta_{3,t} - \mu_3 \end{bmatrix} + \begin{bmatrix} \eta_{1,t+1} \\ \eta_{2,t+1} \\ \eta_{3,t+1} \end{bmatrix}, \tag{6}$$

for $t = 1, \ldots, T$. Next, the measurement equation is given as

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \\ \vdots \\ y_{t,N} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix} \begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \vdots \\ \varepsilon_{t,N} \end{bmatrix}, \tag{7}$$

for $t = 1, \ldots, T$. After rewriting (6) and using a vector/matrix notation, (6) and (7) form the following state space system

$$y_t = \Lambda(\lambda)\beta_t + \varepsilon_t, \tag{8}$$

$$\beta_{t+1} = (I_3 - \Phi)\mu + \Phi\beta_t + \eta_t, \tag{9}$$

for $t = 1, \ldots, T$ and where y_t is the $N \times 1$ yield vector, $\Lambda(\lambda)$ is the loading matrix of size $N \times 3$ and depends only on λ, β_t is the 3×1 state vector, μ is the 3×1 mean vector and Φ is a 3×3 diagonal coefficient matrix. Furthermore, we assume that the measurement and state errors are normally distributed and mutually uncorrelated, i.e.

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{N\times 1} \\ 0_{3\times 1} \end{bmatrix}, \begin{bmatrix} \Sigma_\varepsilon & 0_{N\times 3} \\ 0_{3\times N} & \Sigma_\eta \end{bmatrix} \right), \tag{10}$$

where Σ_ε and Σ_η have dimensions $N \times N$ and 3×3, respectively. Furthermore, we restrict both variance matrices to be diagonal. The restrictions on the variance matrices imply that the error terms of the yields as well as the error terms of the factors are uncorrelated. Diebold et al. (2006) use a VAR(1) structure for the latent factors and do not restrict Σ_η to be diagonal, which allows the state variables to interact dynamically and their shocks to be correlated. However, J. H. Christensen, Diebold, and Rudebusch (2011) perform an out-of-sample forecast for the restricted and unrestricted model and conclude that the forecast performance of the restricted model is better. In this study we are particularly interested in out-of-sample performance. Therefore, we choose to use the restricted DNS model during this study.

As Diebold et al. (2006) point out, the state space format provides a very convenient framework for estimating dynamic factor models. Furthermore, it allows us to estimate the loading parameter alongside the other parameters instead of taking a predefined fixed value. We refer to this model where the loading parameter is estimates alongside the other parameters as the standard DNS model and we use it as a benchmark for the extended models described the upcoming sections. The maximum likelihood estimates of this particular model can be obtained with use of the Kalman filter.

2.3 Time-Varying Loading Parameter

The loading parameter determines the exponential decay of the loading components in the measurement equation and thereby emphatically determines the shape of the yield curve. In many applications of the Nelson-Siegel model this parameter is kept constant. For example, Diebold and Li (2006) fix $\lambda = 0.0609$, which corresponds to a medium term maturity at which the curvature factor achieves its maximum for their dataset. As earlier mentioned it is possible to estimate λ alongside the other parameters with the state space approach. Diebold et al. (2006) do this and find $\lambda = 0.077$, while several others adopt this value for their studies. Diebold and Li (2006) argue that fixing λ results in a simpler estimation and more reliable results. However, Koopman et al. (2010) argue that the characteristics of the yield curve can vary over time such that the maturity for which the loading on the medium term factor achieves its maximum, also changes over time. As the loadings only depend on λ, more flexibility in this parameter may be of great importance. Introducing a time-varying loading parameter results in a non-linear model which makes estimation slightly more complicated. However, the loading parameter is still quite easy to estimate in a state space form. For this reason, they allow for a time-varying loading parameter and obtain results that support their arguments. Because of their encouraging results, we follow their approach.

The time-varying loading parameter λ_t can be included in the set of factors, such that it is considered as a latent factor. However, doing this could lead to negative values for λ_t. Therefore we choose to include $\log(\lambda_t)$ in the set of factors, such that we obtain the new state vector $(\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \log(\lambda_t))'$. We refer to this state vector as α_t. We model the new vector α_t then again using AR(1) processes described by (3). As earlier mentioned, the measurement equation now becomes non-linear in α_t. In this case, we obtain the adjusted state space representation

$$
\begin{aligned}
y_t &= \Lambda(\exp(\alpha_{4,t}))\,(\alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t})' + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \Sigma_\varepsilon\right), \\
\alpha_t &= \left(I_4 - \Phi^L\right)\mu^L + \Phi^L \alpha_t + \eta_t^L, \qquad \eta_t \sim \mathcal{N}\left(0, \Sigma_\eta^L\right),
\end{aligned} \tag{11}
$$

for $t = 1, \ldots, T$ and where the dimension of Φ, μ, η_{t+1} and Σ_η are increased as appropriate. We refer to these matrices/vectors as $\Phi^L, \mu^L, \eta_{t+1}^L$ and Σ_η^L, and we refer to this model as the DNS-L model. An important difference with the representation in (8) and (9) can be found in the matrix Λ, which is now time-varying due to the time-varying loading parameter λ_t. Furthermore, because of the nonlinearity in the measurement equation we can no longer rely on the Kalman filter for estimation. A solution to this problem can be found in estimation using the extended Kalman filter (EKF), which we describe in section 2.8.

2.4 Time-Varying Volatility

Another extension to the DNS model can be found in the allowance for time-varying volatility. Interest rates are known to be related to trading in financial markets. As volatility changes in these markets, the volatility in yields may also change over time. In the DNS model we assume that volatility is constant which may therefore be an inappropriate assumption. Hence, relaxing this assumption possibly results in more accurate estimation of the parameters and better fit of the yield curve. Therefore it seems an relevant extension to examine.

We modify the DNS model by following Koopman et al. (2010), who introduce a common variance component that is modeled as a generalized autoregressive conditional heteroscedasticity (GARCH) process. More specifically, in this approach the disturbances in the measurement equation ε_t are defined as

$$
\varepsilon_t = \Gamma_\varepsilon \varepsilon_t^* + \varepsilon_t^+, \qquad t = 1, \ldots, T, \tag{12}
$$

where Γ_ε and ε_t^+ are vectors of size $N \times 1$ and ε_t^* is a scalar. In this specification, we distinguish between the common disturbance term ε_t^* and the disturbance vector ε_t^+. The disturbance vector is distributed as $\varepsilon_t^+ \sim \text{NID}(0, \Sigma_\varepsilon^+)$, with Σ_ε^+ a $N \times N$ diagonal matrix. For the common disturbance component we assume $\varepsilon_t^* \sim \text{NID}(0, h_t)$. Next, we specify h_t as the GARCH process introduced by Bollerslev (1987). In particular we have

$$
h_t = \gamma_0 + \gamma_1 {\varepsilon_{t-1}^*}^2 + \gamma_2 h_{t-1}, \qquad t = 1, \ldots, T \tag{13}
$$

with coefficients to be estimated $\gamma_0 > 0$, $0 < \gamma_1 < 1$, $0 < \gamma_2 < 1$ and $h_1 = \frac{\gamma_0}{1-\gamma_1-\gamma_2}$. The result is a time-varying variance matrix for ε_t, given by

$$\Sigma_\varepsilon(h_t) = h_t \Gamma_\varepsilon \Gamma_\varepsilon' + \Sigma_\varepsilon^+, \qquad t = 1, \dots, T \tag{14}$$

where $\Sigma_\varepsilon(h_t)$ is a time-varying diagonal variance matrix, whose variation is determined by the univariate common GARCH component h_t.

It should be noticed that the above described setting encounters identification problems. To overcome these problems a restriction is required. Koopman et al. (2010) propose the option to normalize Γ_ε such that $\Gamma_\varepsilon'\Gamma_\varepsilon = 1$ but choose for the option to fix γ_0 at a small value. Here, we follow their choice once again by applying the restriction on γ_0.

To accommodate the property of time-varying volatility in a state space form, we have to make several adjustments to the state space representation of the standard DNS model presented by (8) and (9). In particular, we substitute the new specification for ε_t in the measurement equation. Then, the measurement and its covariance depend on the unobserved common disturbance term ε_t^*. Therefore, ε_t^* needs to be modeled as a latent factor. Hence, we include ε_t^* in the state vector alongside the Nelson-Siegel factors, such that the new state vector equals $(\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \varepsilon_t^*)'$. Furthermore, the common disturbance term does not depend on past factor observations such that its corresponding elements in the state coefficient matrix are all equal to zero. The resulting state space representation of the DNS model with time-varying volatility is given by

$$\begin{aligned} y_t &= \begin{bmatrix} \Lambda(\lambda) & \Gamma_\varepsilon \end{bmatrix} \begin{bmatrix} \beta_t \\ \varepsilon_t^* \end{bmatrix} + \varepsilon_t^+, \quad \varepsilon_t^+ \sim \mathcal{N}(0, \Sigma_{\varepsilon^+}), \\ \begin{bmatrix} \beta_{t+1} \\ \varepsilon_{t+1}^* \end{bmatrix} &= \begin{bmatrix} (I_3 - \Phi)\mu \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi & 0_{3\times1} \\ 0_{1\times3} & 0 \end{bmatrix} \begin{bmatrix} \beta_t \\ \varepsilon_t^* \end{bmatrix} + \begin{bmatrix} \eta_t \\ \varepsilon_{t+1}^* \end{bmatrix}, \quad \begin{bmatrix} \eta_t \\ \varepsilon_{t+1}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_\eta & 0_{3\times1} \\ 0_{1\times3} & h_{t+1} \end{bmatrix} \right), \end{aligned} \tag{15}$$

for $t = 1, \dots, T$. We refer to this model as the DNS-V model. Finally, Koopman et al. (2010) propose a second approach of incorporating varying volatility in the DNS model. In this approach, η_t is decomposed in the same way as (12) and thus the volatility component is included in the state equation instead of the measurement equation. In this way, the volatility component does affect the factors and thereby indirectly affect the yields. However, the volatility of y_t is in this approach affected by the loading matrix $\Lambda(\lambda)$. This can be seen as a restriction to the model described above. In addition, Koopman et al. (2010) find that the maximum likelihood value for this model is substantially lower, which makes this approach less favorable.

2.5 Time-Varying Unconditional Mean

In the standard DNS model (and the extensions mentioned in earlier sections), it is assumed that the factors follow a stationary process with a constant unconditional mean. Dijk et al. (2014) show however that the means in the factor estimates show substantial variation over time and argue that a constant unconditional mean in the factor process may not be appropriate. Therefore, we follow there suggestion and extend the DNS models once more by allowing for a time-varying unconditional mean. We do this by considering two different approaches. The first approach follows Dijk et al. (2014) and considers a purely statistical approach using exponential smoothing, while the second approach follows Exterkate et al. (2013) considering a factor augmented DNS model using macroeconomic information. In the remainder of this section we describe both approaches.

2.5.1 Exponential Smoothing

A straight forward approach to accommodate time variation in the unconditional mean of the factors is to modify (3) to

$$\beta_{j,t+1} = \mu_{j,t+1} + \phi_j(\beta_{j,t} - \mu_{j,t}) + \eta_{j,t+1}, \qquad t = 1, \ldots, T, \tag{16}$$

such that the mean is no longer constant. Now, $\mu_{j,t+1}$ can take various expressions that describe the unconditional mean over time. A simple and intuitively appeiling idea is to let $\mu_{j,t+1}$ depend on its past value and update it with new factor observations. In other words, let $\mu_{j,t+1}$ be a weighted average of $\mu_{j,t}$ and $\beta_{j,t}$. Dijk et al. (2014) implement this idea and come up with a specification where $\mu_{j,t+1}$ is generated by the following exponential smoothing recursion

$$\mu_{j,t+1} = \alpha\beta_{j,t} + (1-\alpha)\,\mu_{j,t}, \qquad t = 1, \ldots, T. \tag{17}$$

In this recursion the decay parameter α takes a value between 0 and 1 and the recursion is started with $\mu_{j,1} = \beta_{j,1}$. If we work this recursion out, we obtain

$$\mu_{j,t+1} = \alpha \sum_{k=0}^{t-2} (1-\alpha)^k \, \beta_{j,t-1} + (1-\alpha)^{t-1} \, \beta_{j,1}, \qquad t = 1, \ldots, T,$$

which shows that $\mu_{j,t+1}$ is an exponentially weighted average of the past factor values. Moreover, if we substitute (17) in (16) we obtain

$$\begin{aligned} \beta_{j,t+1} &= \alpha\beta_{k,t} + (1-\alpha)\,\mu_{j,t} + \phi_j\beta_{j,t} - \phi\mu_{j,t} + \eta_{j,t+1} \\ &= (\alpha + \phi_j)\,\beta_{j,t} + (1 - \alpha - \phi_j)\,\mu_{j,t} + \eta_{j,t+1} \\ &= w_j\beta_{j,t} + (1-w_j)\,\mu_{j,t} + \eta_{j,t+1}, \end{aligned} \tag{18}$$

for $t = 1, \ldots, T$ and $w_j = \alpha + \phi_j$. From this, it follows that the conditional expectation of the factor at time $t+1$ is an exponential weighted average of the factor realization at time t and the unconditional expectation. The smoothness of the unconditional mean is determined by the the decay parameter α. The closer α gets to zero, the smoother the forecasts become. Dijk et al. (2014) set $\alpha = 0.1$ for their analysis of monthly series. Here, we estimate α alongside with the other parameters.

To introduce a time-varying unconditional mean in the state space form, we need to make some adjustments to the structure in the state equation as well as the measurement equation of the standard DNS model. First, we expand the state variable by including μ_t, such that it becomes $x_t = (\beta_t, \mu_t)'$. Since the elements of μ_t are not included in the measurement equation, we expand the measurement coefficient matrix by adding three zero vectors of length N next to $\Lambda(\lambda)$. Then, we use (17) and (18) to form the following state equation

$$
\begin{aligned}
y_t &= \begin{bmatrix} \Lambda(\lambda) & 0_{N\times 3} \end{bmatrix} \begin{bmatrix} \beta_t \\ \mu_t \end{bmatrix} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon) \\
\begin{bmatrix} \beta_{t+1} \\ \mu_{t+1} \end{bmatrix} &= \begin{bmatrix} \alpha I_3 + \Phi & (1-\alpha) I_3 - \Phi \\ \alpha I_3 & (1-\alpha) I_3 \end{bmatrix} \begin{bmatrix} \beta_t \\ \mu_t \end{bmatrix} + \eta_t^{ES}, \quad \eta_t^{ES} \sim \mathcal{N}\left(0, \Sigma_\eta^{ES}\right),
\end{aligned} \tag{19}
$$

for $t = 1, \ldots, T$, where the dimension of η_t and Σ_η changed as appropriate. We refer to this vector and matrix as η_t^{ES} and Σ_η^{ES}. Moreover, the specification for μ_{t+1} is not stochastic. Hence, it does not contain any uncertainty and the disturbance vector and the state variance matrix take the following structures

$$
\eta_t^{ES} = \begin{bmatrix} \eta_t \\ 0_{3\times 1} \end{bmatrix}, \quad \Sigma_\eta^{ES} = \begin{bmatrix} \Sigma_\eta & 0_{3\times 3} \\ 0_{3\times 3} & 0_{3\times 3} \end{bmatrix},
$$

where $0_{3\times 3}$ is a zero matrix of size 3×3. We refer to this model as the DNS-M(ES) model.

2.5.2 Macroeconomic Information

Until this moment we only considered models that solely make use of yield data, ignoring any link to the macroeconomy. This is a possible shortcoming of these models since it is well known that the dynamics of the yield curve are related to the macroeconomy. In particular, important economic factors as inflation, productivity and the federal fund rate are often seen as drivers for the yields curve. Hence, factors like these should have potential to improve the performance of models when they are used. For this purpose, Diebold et al. (2006) include measures for inflation, production capacity utilization and the federal funds rate in the standard DNS model. Others, like De Pooter, Ravazzolo, and Van Dijk (2010) and Exterkate et al. (2013) consider factor construction methods to exploit information of a large set of variables. Diebold et al. (2006) describe an approach that includes the macroeconomic factors into the factor specifications. The resulting model is often referred to as the Factor Augmented Nelson-Siegel (FADNS) model. Consider the AR(1) specification for the factors in the DNS model

$$
\beta_{j,t} = c_j + \phi_j \beta_{j,t-1} + \eta_t, \qquad t = 1 \ldots, T, \tag{20}
$$

where $c = (1-\phi_j)\mu_i$. In this equation we know that if $|\phi_j| < 1$, the unconditional mean equals $c/(1-\phi_j)$. When we include p macroeconomic factors $\{f_{i,t}\}_{i=1}^p$ to the factor specification we obtain the following new specification

$$\beta_{j,t} = c_{j,t} + \phi_j \beta_{j,t-1} + \eta_t, \qquad t = 1, \dots, T, \tag{21}$$

where $c_t = (1 - \phi_j)\mu + a_1 f_1 + a_2 f_2 + \ldots + a_p f_p$. In this specification the parameter μ loses its interpretation of the unconditional mean. The difference with (20) lies obviously in the time-varying intercept $c_{j,t}$. The time-varying intercept implies a time-varying unconditional mean. Hence, we can consider the FADNS model as an approach to introduce a time-varying unconditional mean in the DNS model. In this section we first discuss how we include additional factors into the DNS model and then describe two approaches for constructing the factors to be included. In section 4.4, we discuss the number of factors that we include.

Factor-augmented Nelson-Siegel model

For now we assume that we have p factors available that represent information of the macroeconomy. We include these factors in the state vector such that it becomes $(\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, f_{1,t}, \dots, f_{p,t})'$. With the extension of the state factor, the size of the coefficient matrix in the state equation increases from 3×3 to $(3+p)\times(3+p)$. Hence, the number of parameters to be estimated rapidly increases. For example, including three macro factors would result in $36 - 9 = 27$ additional parameters in the state transition matrix only. This is not desirable because of resulting parameter uncertainty. For this reason Exterkate et al. (2013) apply restrictions to the model and find that it performs better than the unrestricted model. We therefore follow their suggestions for putting restricting to the model and restrict the state transition matrix to adopt the following structure

$$\Phi^{FA} = \begin{bmatrix} \text{diagonal} & \text{unrestricted} \\ \text{zero} & \text{diagonal} \end{bmatrix},$$

where the blocking represents the partitioning of β_t and f_t. This structure implies that the macro factors do not depend on the Nelson-Siegel factors nor other macro factors and there is no interaction between the Nelson-Siegel factors. Furthermore, we restrict both covariance matrices Q and R again to be diagonal. The resulting state space form is then given by

$$\begin{aligned} y_t &= \begin{bmatrix} \Lambda(\lambda) & 0_{N\times p} \end{bmatrix} \begin{bmatrix} \beta_t \\ f_t \end{bmatrix} + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \Sigma_\varepsilon), \\ \begin{bmatrix} \beta_{t+1} \\ f_{t+1} \end{bmatrix} &= \left(I_{3+p} - \Phi^{FA}\right) \begin{bmatrix} \mu \\ 0_{p\times1} \end{bmatrix} + \Phi^{FA} \begin{bmatrix} \beta_t \\ f_t \end{bmatrix} + \eta_t^{FA} \qquad \eta_t^{FA} \sim \text{NID}\left(0_{3+p}, \Sigma_\eta^{FA}\right), \end{aligned} \tag{22}$$

for $t = 1, \dots, T$, where the dimensions of Φ, η_{t+1} and Σ_η are increased as appropriate and Φ is restricted as mentioned above. We refer to these matrices/vectors as $\Phi^{FA}, \eta_{t+1}^{FA}$ and Σ_η^{FA}, and we refer to this model as the DNS-M(PCA) or the DNS-M(PLS) model depending on the factor construction method.

Factor extraction using principal component analysis

Currently there is an enormous availability of macroeconomic data and there are simply way too many macroeconomic variables to include them all in the DNS model. Doing this would lead to very inaccurate estimates due to overfitting, which is of course undesirable. Therefore, one wants to find a way of including the most important information of a large macro set by using a small number of factors. One way to do this is by applying pre-selection methods to pick a few variables that represent the macro set the best. For example, Diebold et al. (2006) choose to use measures for inflation, the real economy and the monetary policy. However, one can argue that besides these variables many other variables may have effect on the evolution of the yield curve. By selecting a limited number of variables one omits potential important information available in other macroeconomic variables. Since there is a large amount of macroeconomic variables available, we are interested in exploiting information from as much macro variables as possible without incorporating a high level of parameter uncertainty and chance of overfitting in the estimation. A possible solution can be found in methods that summarize the information from a large panel of macro variables into a limited number of factors. Mönch (2008) and De Pooter (2007) include a small number of principal components for this purpose.

The idea of PCA is to describe the variation in a large set of variables with a limited number of factors, where the factors are linear combinations of the considered variables. These factors should optimally explain the variation in the variables. Principal components turn out to do this. For the construction of the principal components we denote M as the set of k macroeconomic variables and calculate the corresponding covariance matrix. Next, we calculate the eigenvalues $\{v_j\}_{j=1}^k$ and the eigenvectors $\{e_j\}_{j=1}^k$ of the covariance matrix and sort the eigenvectors in descending order based on their associated eigenvalue. The resulting matrix is denoted as E. We can calculate the matrix of principal components P now by

$$P = ME. \tag{23}$$

The j-th principal component p_j is then defined as the j-th column of P, which can be written as

$$p_j = e_{i,1}m_1 + e_{j,2}m_2 + \ldots + e_{j,q}m_q, \qquad j = 1, \ldots, k, \tag{24}$$

where m_j is the j-th column of M and equals the j-th macro variable in the set of macro variables. Furthermore, the proportion of the total variation that is explained by the j-th principal component is equal to $v_j / \sum_{j=1}^k v_j$.

Factor extraction using partial least squares

Although PCA is a simple and convenient approach for reducing the dimension of a set of variables, it exhibits the drawback that it only captures characteristics of the predictive variables. Partial least squares (PLS) is a factor construction method that finds orthogonal linear combinations of variables with explanatory power for a set of dependent variables. The difference with PCA lies in the fact that PLS constructs factors to explain the variance of the set of dependent variables whereas PCA constructs factors to explain the variance in the explanatory variables. Thus, in our setting of the DNS model, PLS takes the explanatory power of the macroeconomic variables for the Nelson-Siegel factors into account with the construction of the factors. Following Garthwaite (1994) we describe multivariate partial least squares as sequential regressions.

Consider the set of explanatory variables $M = (m_1, m_2, \dots, m_k)$ and let $Z = (z_1, z_2, \dots, z_q)$ be a set of response variables. The aim of PLS is to find a set of factors $F = (f_1, f_2, \dots, f_p)$, with $p \ll k$, that explains all the variables in Z well by constructing the factors as linear combinations of the variables in M. To construct the first factor f_1 we define $g_{1,s} = z_s - \bar{z}_s$ and $d_{1,j} = m_j - \bar{m}_j$, and the matrices $G_1 = (g_{1,1}, g_{1,2}, \dots, g_{1,q})$ and $D_1 = (d_{1,1}, d_{1,2}, \dots, d_{1,k})$. Next, we construct the eigenvector corresponding to the largest eigenvalue of $G_1' D_1 D_1' G_1$ and denote it by c_1. Then, we define $u_1 = G_1 c_1$ and perform regressions of u_1 on $d_{1,j}$ for $j = 1, 2, \dots, k$, such that we obtain k estimates $\hat{u}_{1,j} = b_{1,j} d_{1,j}$, where $b_{1,j} = (d_{1,j}' d_{1,j})^{-1} d_{1,j}' u_1$. The first factor is then obtained by taking a weighted average of the estimates $\hat{U}_1 = (\hat{u}_{1,1}, \hat{u}_{1,2}, \dots, \hat{u}_{1,k})$, i.e.

$$f_1 = \sum_{j=1}^{k} w_j \hat{u}_{1,j} = \sum_{j=1}^{k} w_j b_{1,j} d_{1,j}. \tag{25}$$

The used weights w_j in this specification are discussed in a later part of this subsection. Although f_1 is a weighted average of the predictors in u_1 and is a useful predictor for Z, the variables in M potentially contain further useful information for explaining Z. The information in M that has not been exploited using the first factor can be estimated by the residuals of regressions of $d_{1,j}$ on f_1, for $j = 1, \dots, k$. Similarly, the unexplained part in Z can be estimated by the residuals of regressions of $g_{1,s}$ on f_1, for $s = 1, \dots, q$. These residuals are denoted as $d_{2,j}$ for $d_{1,j}$ and $g_{2,s}$ for $g_{1,s}$. Next, we let c_2 be the eigenvector corresponding to the largest eigenvector of $G_2' D_2 D_2' G_2$ and compute u_2 as $u_2 = G_2 c_2$. Then, we can construct a new factor f_2 that tries to capture the remaining information in Z in the same way as described above, but now using u_2 and D_2. Continuing this way we can construct $f_2, \dots, f_p$ iteratively, where each factor depends on regressions on the previous factor. In general form, constructing f_i comes down to the following procedure. Assume that we have computed f_i and $d_{i,j}$ for $j = 1, \dots, k$ and $g_{i,s}$ for $s = 1, \dots, q$. Then, $d_{i+1,j}$ and $g_{i+1,s}$ are the residuals from the regressions of $d_{i,j}$ on f_i and $g_{i,s}$ on f_i, respectively. Hence, we obtain

$$d_{i+1,j} = d_{i,j} - (f_i' f_i)^{-1} f_i' d_{i,j} \cdot f_i \qquad \text{and} \qquad g_{i+1,s} = g_{i,s} - (f_i' f_i)^{-1} f_i' g_{i,s} \cdot f_i,$$

for $j = 1, \dots, k$ and $s = 1, \dots, q$. Next, we let c_{i+1} be the eigenvector corresponding to the largest eigenvalue of $G_{i+1}' D_{i+1} D_{i+1}' G_{i+1}$ and denote

$$u_{i+1} = G_{i+1} c_{i+1}.$$

In the final stage, we now obtain k estimates for u_{i+1} by regressing u_{i+1} on $d_{i+1,j}$, i.e.

$$\hat{u}_{i+1,j} = b_{i+1,j} d_{i+i,j} \quad \text{for } j = 1, \dots, k, \tag{26}$$

with $b_{i+1,j} = (d'_{i+1,j} d_{i+1,j})^{-1} d'_{i+1,j} u_{i+1}$. By creating a linear combination of these predictors, we now construct the $(i+1)$-th factor as

$$f_{i+1} = \sum_{j=1}^{m} w_j \hat{u}_{i+1,j} = \sum_{j=1}^{m} w_{i+1,j} b_{i+1,j} d_{i+1,j}. \tag{27}$$

To complete the procedure, we specify the weights as $w_{i,j} = d'_{i,j} d_{i,j}$ such that they are proportional to the variance of the columns in D_i. Using this weighting scheme, the linear combination of the vectors in D_i depends on the covariance with the variable we aim to explain.

Since we include the macroeconomic factors in the state equation, the objectives that we intend to explain are the Nelson-Siegel factors. However, a part of the factors will already be explained by the autoregressive term in the specification. Consequently, we are particularly interested in whether the macroeconomic factors can explain the part that is not explained by the autoregressive term. Therefore, we let the response variables in the above described procedure be the residuals of the Nelson-Siegel factors after applying the 2-step estimation procedure described in section 2.2.

2.6 Combination of Extensions

The extensions mentioned in sections 2.3 and 2.4 do all have the potential to improve in-sample fit as well as out-of-sample forecasts compared to the standard DNS model. Applying these extensions all in one advanced model could therefore improve performance even further. Since all extensions are clearly covered in previous sections, we only discuss the state space representation of the DNS model that combines the extensions of a time-varying loading parameter, time-varying volatility and a time-varying unconditional mean. We refer to this model as the DNS-LVM model. Furthermore, we distinguish between the DNS-LVM model using exponential smoothing (DNS-LVM(ES)) and the DNS-LVM model using the factor augmented structure (DNS-LVM(FA)) because both require different modifications to the state space representation. For the DNS-LVM(ES) model, the loading parameter λ_t and the common disturbance term ε_t^* and the five unconditional means are included in the state vector such that the state vector becomes $\{\beta_{t,1}, \beta_{t,2}, \beta_{t,3}, \lambda_t, \varepsilon_t^*, \mu_{t,1}, \cdots, \mu_{t,5}\}$. The state space representation is given by

$$\begin{gathered} y_t = \Lambda\left(\exp(\alpha_{4,t})\right)(\alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t})' + \Gamma\varepsilon_t^* + \varepsilon_t^+, \quad \varepsilon_t^+ \sim \mathcal{N}\left(0, \Sigma_\varepsilon\right), \\ \begin{bmatrix} \alpha_{t+1} \\ \varepsilon_{t+1}^* \\ \mu_{t+1} \end{bmatrix} = \begin{bmatrix} aI_5 + \Phi^{LV} & (1-a)I_5 - \Phi^{LV} \\ aI_5 & (1-a)I_5 \end{bmatrix} \begin{bmatrix} \alpha_t \\ \varepsilon_t^* \\ \mu_t \end{bmatrix} + \nu_{t+1}, \quad \nu_{t+1} \sim \mathcal{N}\left(0, \Sigma_\nu^{LVME}\right), \end{gathered} \tag{28}$$

for $t = 1, \ldots, T$, where

$$\Phi^{LV} = \begin{bmatrix} \Phi^L & 0_{4\times1} \\ 0_{1\times4} & 0 \end{bmatrix}, \quad \nu_{t+1} = \begin{bmatrix} \eta^L_{t+1} \\ \varepsilon^*_{t+1} \\ 0_{5\times1} \end{bmatrix}, \quad \Sigma^{LVME}_{\nu} = \begin{bmatrix} \Sigma^L_\eta & 0_{4\times1} & 0_{4\times5} \\ 0_{1\times4} & h_t & 0_{1\times5} \\ 0_{5\times4} & 0_{5\times1} & 0_{5\times5} \end{bmatrix}$$

and $\mu_t = (\mu_{t,1}, \cdots, \mu_{t,5})'$. For the DNS-LVM(FA) model, we include besides the loading parameter and the common shock component, also the factors containing macroeconomic information. These factors are either constructed by PCA or PLS. Hence, the new state vector becomes $(\beta_{t,1}, \beta_{t,2}, \beta_{t,3}, \lambda_t, \varepsilon^*_t, f_{t,1}, \cdots, f_{t,p})'$. The corresponding state space representation is now given by

$$y_t = \Lambda\left(\exp\left(\alpha_{4,t}\right)\right)\left(\alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t}\right)' + \Gamma\varepsilon^*_t + \varepsilon^+_t, \quad \varepsilon^+_t \sim \mathcal{N}\left(0, \Sigma_{\varepsilon^+}\right),$$

$$\begin{bmatrix} \alpha_{t+1} \\ f_{t+1} \\ \varepsilon^*_{t+1} \end{bmatrix} = \begin{bmatrix} \left(I_4 - \Phi^{LVMF}\right) & 0_{(4+p)\times1} \\ 0_{1\times(4+p)} & 0 \end{bmatrix} \begin{bmatrix} \mu^{LVMF} \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi^{LVMF} & 0_{(4+p)\times1} \\ 0_{1\times(4+p)} & 0 \end{bmatrix} \begin{bmatrix} \alpha_t \\ f_t \\ \varepsilon^*_t \end{bmatrix} + \begin{bmatrix} \eta^{LVMF}_t \\ \varepsilon^*_{t+1} \end{bmatrix}, \tag{29}$$

for $t = 1, \ldots, T$, where

$$\begin{bmatrix} \eta^{LVMF}_t \\ \varepsilon^*_{t+1} \end{bmatrix} \sim \mathcal{N}\left(0_{(5+p)\times1}, \begin{bmatrix} \Sigma^{LVMF}_\eta & 0_{(4+p)\times1} \\ 0_{1\times(4+p)} & h_{t+1} \end{bmatrix}\right).$$

The dimensions of Φ, μ, η_{t+1} and Σ_η have increased appropriately and are denoted by Φ^{LVMF}, μ^{LVMF} and η^{LVMF}_{t+1}, respectively. Finally, Φ^{LVMF} and μ^{LVMF} adopt the following structures

$$\Phi^{LVMF} = \begin{bmatrix} \text{diagonal} & \text{unrestricted} & 0_{4\times1} \\ \text{zero} & \text{diagonal} & 0_{p\times1} \\ 0_{1\times4} & 0_{1\times p} & 0 \end{bmatrix}, \quad \mu_{t+1} = \begin{bmatrix} \mu^L \\ 0_{(p+1)\times1} \end{bmatrix},$$

where the blocking in Φ^{LVMF} corresponds to the partitioning of the state vector into α_t and f_t. The interpretation of this structure is almost equal to that of the FADNS models described in subsection 2.5.2. That is, we do not model feedback from the factors in α_t to the macro factors nor dynamic interaction between the factors in α_t and between the macro factors.

2.7 Estimation of Linear State Space Models

The DNS models specified in subsections 2.2 - 2.6 contain latent variables that need to be estimated. The Kalman filter is a recursive procedure that uses the information from the data at time $t-1$ to construct an optimal estimate of the latent factor at time t.

Consider the general state space representation in (4) and (5). Besides the latent factors, the parameters in A, C, H, Q and R are also unknown and need to be estimated as well.

However, the Kalman filter requires the parameters to be known to construct estimates for the latent factors. Therefore, we assume for now that the parameters are known. For readers that are interested in the derivation of the Kalman filter I refer to appendix C, here I only present the so called filtering- and prediction steps for the reason of brevity. For given values of $x_{t|t-1}$ and $P_{t|t-1}$ the filtering step is given by

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t v_t, \tag{30}$$

$$P_{t|t} = P_{t|t-1} - K_t H' P_{t|t-1}, \tag{31}$$

where $v_t = y_t - H'\hat{x}_{t|t-1}$ is the prediction error, $K_t = P_{t|t-1} H' F_t^{-1}$ is called the Kalman gain and $F_t = H P_{t|t-1} H' + R$ is the measurement prediction variance. Using the state equation in (5), the prediction step is then given by

$$\hat{x}_{t+1|t} = C + A\hat{x}_{t|t}, \tag{32}$$

$$P_{t+1|t} = A P_{t|t} A' + Q. \tag{33}$$

Finally, we initialize the recursion by using the unconditional mean and variance of x_t, i.e.

$$\hat{x}_{1|0} = \mathrm{E}\left[x_t\right] = \mu, \qquad P_{1|0} = \mathrm{E}\left[x_t x_t'\right] = \Sigma_x,$$

where the unconditional variance can be obtained by solving $\Sigma_x - A\Sigma_x A' = \Sigma_\nu$. The procedure to solve this equation is described in appendix B.

The Kalman filter is able to provide minimum mean square linear estimates for the latent factors, conditional on the parameters in (4) and (5). However, these parameters are unknown in the DNS model and need to be estimated. For this purpose we use maximum likelihood. Let θ be the vector that collects all unknown coefficients in A, C, H, Q and R. The joint pdf $f(y_1, y_2, \cdots, y_T|\theta)$ can be rewritten as the product of conditional pdfs. As we assumed that the errors terms ω_t and ν_t are Gaussian, conditional pdfs are also Gaussian. Hence, it holds that

$$y_t|\mathcal{I}_{t-1} \sim \mathcal{N}\left(H\hat{x}_{t|t-1}, F_t\right),$$

such that the log-likelihood function becomes

$$l(\theta) = \sum_{t=1}^{T} f(y_t|\mathcal{I}_{t-1};\theta) = -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T} v_t' F_t^{-1} v_t. \tag{34}$$

It is possible to evaluate $l(\theta)$ using the Kalman filter. Numerically optimizing $l(\theta)$ results in maximum likelihood estimates of the parameters in θ.[1] The procedure for obtaining the optimal latent factors and parameters is recursive in the sense that the optimal latent factors depend on

[1] In practice, the optimization of the log-likelihood function can encounter problems due to the inversion of F_t. Therefore we choose to use the univariate variant of the Kalman filter. This approach is explained in appendix D.

the parameters and vice versa. To start the procedure, one requires a set of initial parameters. For this purpose we carry out the 2-step procedure as described in section 2.2.

For the models with time-varying volatility, some additional steps are required. In these cases, the variance matrix Q depends on the volatility specification described in (13). This implies a time-varying variance matrix Q_t. Moreover, we need the volatility component h_{t+1} to make a prediction. However, h_{t+1} is a function of the unobserved error term ε_t^*, such that it is not observed. Hence, we are not able to compute the value of h_{t+1} at time t. A solution can be found in taking the expectation of the latent term in the volatility specification such that we obtain an estimate for h_{t+1}. This gives

$$\hat{h}_{t+1|t} = \gamma_0 + \gamma_1 \,\mathrm{E}[\varepsilon_t^{*2}|\mathcal{I}_t] + \gamma_2 \hat{h}_{t|t-1}. \tag{35}$$

To calculate the expectation term $\mathrm{E}[\varepsilon_t^{*2}|\mathcal{I}_t]$ we consider the triviality

$$\varepsilon_t^* = \mathrm{E}[\varepsilon_{t-1}|\mathcal{I}_t] + (\varepsilon_t^* - \mathrm{E}[\varepsilon_t^*|\mathcal{I}_t]).$$

From this it can be easily shown by quadrating and taking conditional expectations that

$$\mathrm{E}[\varepsilon_t^{*2}|\mathcal{I}_t] = \mathrm{E}[\varepsilon_t^*|\mathcal{I}_t]^2 + \mathrm{E}[(\varepsilon_t^* - \mathrm{E}[\varepsilon_t^*|\mathcal{I}_t])^2],$$

where the first term is last element of $\hat{x}_{t|t}$ and the second term is the last diagonal element of $P_{t|t}$. Next, we plug this expression into (35) to obtain a prediction for the volatility component h_{t+1}. Finally, we substitute the predicted value h_{t+1} in the variance matrix Q_t at the row corresponding with the place of ε_t^* in the state vector.

2.8 Estimation of Nonlinear State Space Models

As demonstrated in subsection 2.7, the Kalman filter is able to provide state estimations in a linear state space model. However, assumptions of linearity are essential for the derivation of the Kalman filter. When a nonlinear transformation is applied to the Gaussian random variable, the random variable holds no longer a Gaussian distribution. With the introduction of a time-varying loading parameter, the state space system becomes nonlinear. Hence, in this setting the Kalman filter becomes useless. Luckily we can proceed with an extension on the Kalman filter, which relaxes the assumption of linearity by assuming that the equations include a nonlinear function. Consider the general nonlinear state space representation

$$\begin{aligned} y_t &= Z(x_t) + \omega_t, \qquad \omega \sim \mathcal{N}(0, R), \\ x_t &= C + Ax_t + \eta_t, \qquad \nu_t \sim \mathcal{N}(0, Q), \end{aligned} \tag{36}$$

for $t = 1, \ldots, T$, where $Z(x_t)$ is a nonlinear function of the state variable x_t. As discussed above, the nonlinearity in $Z(x_t)$ implies that the state variable does no longer follow a Gaussian distribution. To circumvent this problem, the extended Kalman filter locally linearizes $Z(x_t)$

at $x_t = \hat{x}_{t|t-1}$, where $\hat{x}_{t|t-1}$ is an estimate of x_t based on the past observations up to time $t-1$. This results in the following approximation

$$Z_t(x_t) \approx Z_t\left(\hat{x}_{t|t-1}\right) + \dot{Z}_t\left(x_t - \hat{x}_{t|t-1}\right), \tag{37}$$

with Jacobian matrix $\dot{Z}_t = \partial Z_t(x_t)/\partial x_t|_{x_t=\hat{x}_{t|t-1}}$. This Jacobian matrix and its derivation is described in appendix F for every non-linear model described in the previous subsections. If we now substitute (37) in the measurement equation of (36) we obtain the following linearized model

$$y_t = d_t + \dot{Z}_t x_t + \omega_t, \tag{38}$$

with $d_t = Z_t\left(\hat{x}_{t|t-1}\right) - \dot{Z}_t\hat{x}_{t|t-1}$. Since we have at this point again a linear system, we can apply the usual filtering step of the Kalman filter, which is now given by

$$\begin{aligned} \hat{x}_{t|t} &= \xi_{t|t-1} + K_t v_t, \\ P_{t|t} &= P_{t|t-1} - K_t \dot{Z}_t P_{t|t-1}, \end{aligned} \tag{39}$$

where $v_t = y_t - Z_t\left(\hat{x}_{t|t-1}\right)$, $K_t = V_{t|t-1}\dot{Z}_t' F_t^{-1}$, with $F_t = \dot{Z}_t V_{t|t-1}\dot{Z}_t' + R$. The next step is to obtain a prediction for the following observation. The prediction step is the same as in (57) and (58), i.e.

$$\begin{aligned} \hat{x}_{t+1|t} &= C + A\hat{x}_{t|t} \\ P_{t+1|t} &= AP_{t|t}A' + Q. \end{aligned} \tag{40}$$

Due to the linearization of the extended Kalman filter the estimates $\hat{x}_{t|t}$ and $P_{t|t}$ are suboptimal, such that we should consider $P_{t|t-1}$ and $P_{t|t}$ as approximate mean squared error matrices.

Finally we obtain parameter estimates by substituting v_t and F_t into the log-likelihood in (34) and numerically optimizing the function in the same manner as with the linear state space model. To start the recursive procedure we obtain starting parameters using the 2-step approach again, but now using nonlinear least squares instead of ordinary least squares.

3 Data

3.1 Yield Data

For our analysis we make use of monthly yield data, that we compose from daily yields provided by the United States Department of the Treasury.[2] The resulting yield dataset consists of monthly zero coupon US government bond yield observations for maturities of 3, 6, 12, 24, 36, 60, 84, 120, 240 and 360 months. Together, this results in a set of yields for ten maturities. The dataset contains end-of-month yields and spans the time period from October 1993 through January 2017, resulting in 280 observations per series.

[2] Available at `https://treasury.gov/resource-center/data-chart-center`.

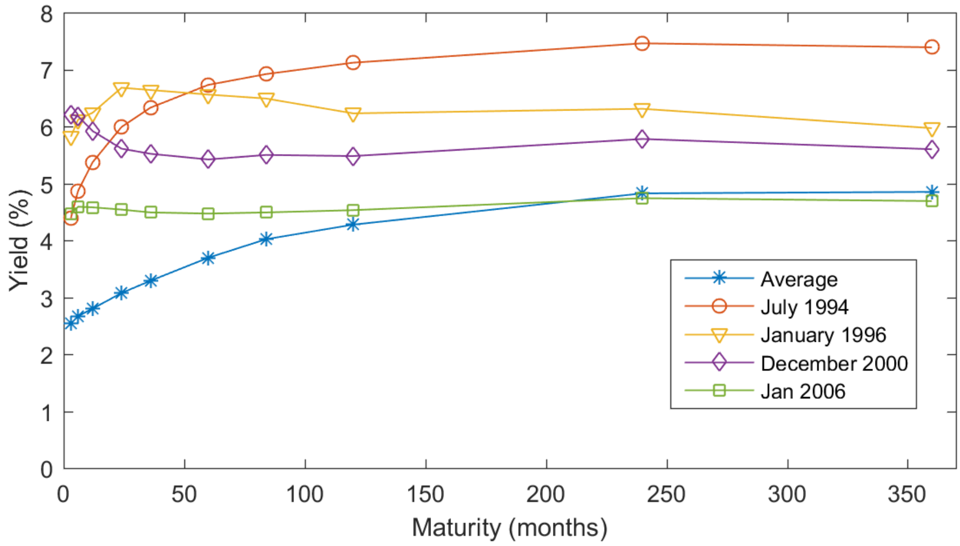


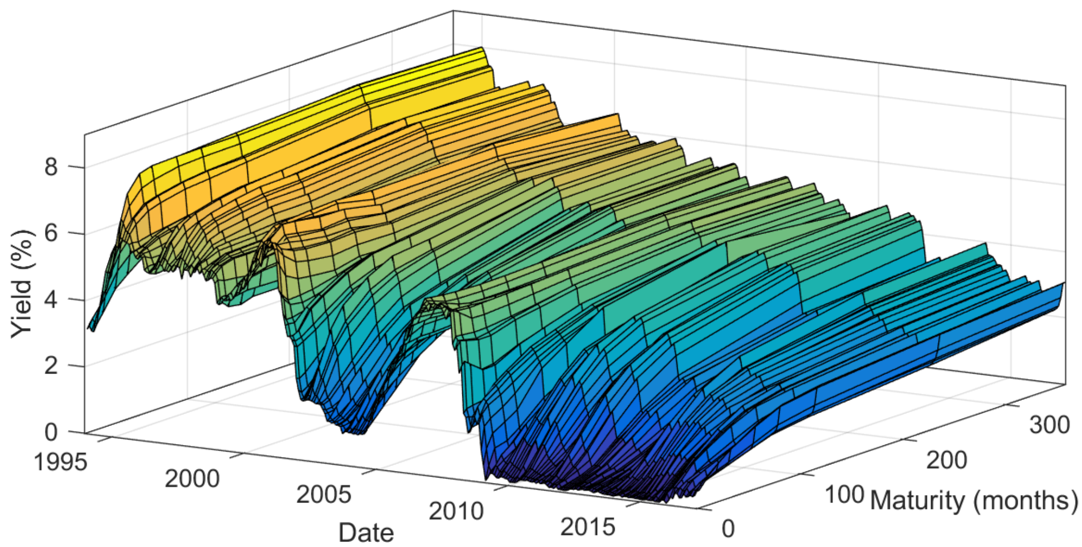
Figure 2: Different shapes of the yield curve. This figure displays the average yield curve over the period from October 1993 through January 2017 and displays the yield curve in four selected months.

Except for the 360-month yield, yields for all maturities are available over the complete time period. The department of treasury ceased the publication of the 360-month yield in February 2002 and reintroduced the series in February 2006. In order to estimate the 360-month yield during this period, the Department of Treasury suggests to use certain extrapolation factors that are provided on the website of the Department of the Treasury. We follow their suggested approach by adding the extrapolation factor to the 240-month yield during the period in which the Department of Treasury stopped selling 360-month bonds. Before and beyond this period we use the regular 360-month yield.

Figure 2 shows that the average yield curve is slowly concavely increasing. This implies that in general a longer maturity results in a higher yield. This makes sense as the lender is subject to higher risk from growing chances the yields and inflation rise. Therefore he expects a higher yield in return. However, the figure also illustrates that a variety of shapes occur over the time period. The typical yield curve is concavely increasing (1994). Other shapes that occur are “humped” (1996), “decreasing” (2000) and “flat” (2006). These shapes are rare compared to the typical yield curve and occur usually in particular situations. A flat yield curve is typically related to uncertainty among investors about the future outlook of the economy, a humped is often related to a slowing economy and a decreasing curve relates to a high probability of a recession. Figure 3a shows a three-dimensional plot of the dataset and confirms the presence of a variety of shapes during the sample period. From this figure, we also observe that yields vary strongly over time. Overall we observe a decreasing pattern of the yields over time, although this is clearer for yields with longer maturities. These finding are supported by the plots of yields for several maturities in figure 3b. Yields with shorter maturities in figure 3b seem to show a more incoherent pattern and deviate more from the overall trend. Hence, we can conclude that volatility decreases when the time to maturity increases.

The concavely increasing shape of the average yield curve implies that in general a longer

(a)



(b)

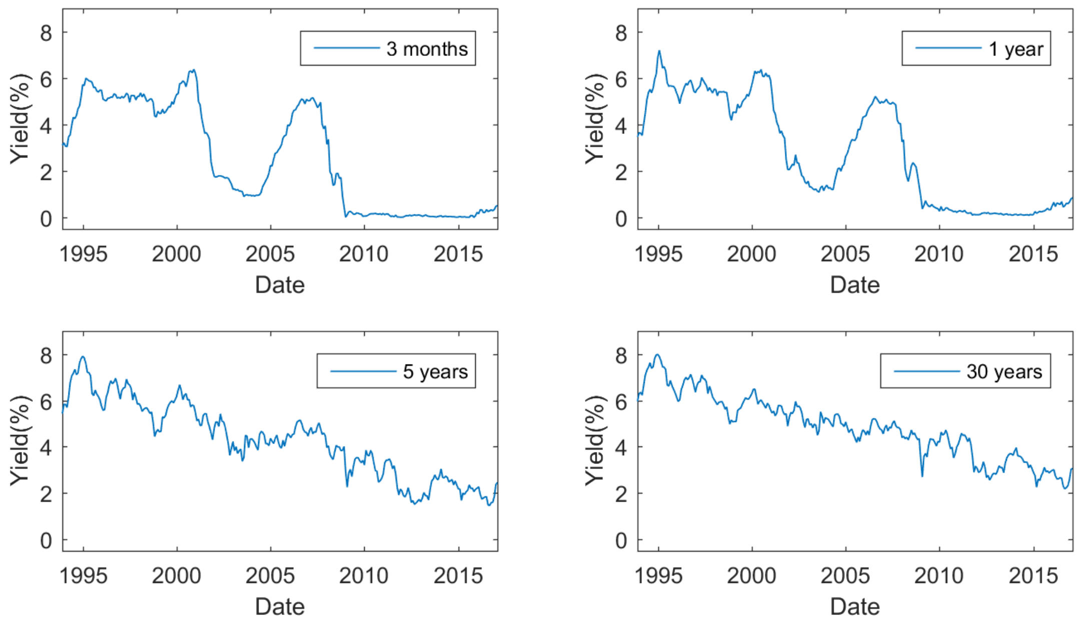


Figure 3: Yield data visualization. This figure shows an overview of the yield curves for the period October 1993 through January 2017. The sample contains monthly data for maturities maturities of 3, 6, 12, 24, 36, 60, 84, 120, 240 and 360 months.

Table 1: Descriptive statistics of yield data

Maturity	Mean	Std	Skew	Kurt	Min	Max	JB	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(36)$
3	2.537	2.242	0.181	1.348	0.000	6.380	33.361	0.994	0.807	0.313
6	2.670	2.282	0.173	1.358	0.030	6.510	32.868	0.994	0.808	0.314
12	2.800	2.284	0.168	1.404	0.090	7.200	31.028	0.994	0.819	0.355
24	3.075	2.258	0.164	1.502	0.200	7.690	27.429	0.992	0.826	0.426
36	3.289	2.165	0.145	1.583	0.300	7.800	24.396	0.991	0.825	0.469
60	3.695	1.958	0.101	1.760	0.590	7.830	18.426	0.988	0.813	0.520
84	4.020	1.805	0.082	1.874	0.980	7.840	15.104	0.986	0.800	0.544
120	4.275	1.634	0.077	2.051	1.460	7.910	10.792	0.984	0.779	0.541
240	4.825	1.537	-0.064	2.161	1.780	8.100	8.397	0.983	0.779	0.554
360 (level)	4.851	1.373	0.065	2.285	2.180	7.990	6.158	0.981	0.757	0.541
Slope	2.313	1.397	-0.298	1.940	-0.640	4.570	17.251	0.978	0.563	-0.291
Curvature	-0.810	1.076	-0.033	2.265	-3.120	2.030	6.356	0.948	0.560	0.260

NOTE: This table reports the descriptive statistics for the monthly US zero-coupon yields at different maturities over the period January 2000 to August 2016. More specific, it reports the mean, standard deviation, skewness, kurtosis, minimum, maximum, the Jarque-Bera statistic and the 1-, 12- and 36-month autocorrelation for the time series of different maturities. The maturities are expressed in months In addition, it reports the same characteristics for empirical proxies of the level, slope and curvature factor. The proxy for the level equals the longest maturity yield (360 months), for the slope factor it is computed as the 30-year yield minus the 3-month yield and for the curvature factor it is computed as two times the 36-month yield minus the 360-month yield and 3-month yield.

maturity results in a higher yield. This makes sense as the lender is subject to higher risk from growing chances the yields and inflation rise. The big difference between the development of short- and long-term yields can be explained by the fact that short-term yields are more sensitive to decisions of the Federal Reserve Board's Open Market Committee (FOMC) on the Federal Funds Rate, whereas long-term yields depend more on supply and demand and expectations of inflation. After the financial crisis in 2008, short-term yields rapidly decreased and stabilized at an extreme low level close to zero until the end of the sample period. This may be regarded as a remarkable development the history of government bond yields.

Table 1 reports the mean, standard deviation, minimum, maximum as well as the 1-month, 1-year and 3-year autocorrelation for yields with the discussed set of maturities and of the proxies for the slope en curvature factors. The proxy for the level factor equals the longest maturity yield (360 months), for the slope it is defined as the 360-month yield minus the 3-month yields and for the curvature it is defined as the 36-month yield minus the 360-month yield and 3-month yield. The mean values for the different maturities confirm the slowly concavely increasing pattern of the average yield curve over time, whereas the standard deviations confirm the larger dispersion in yields over time for shorter maturities. There are no clearly notable differences across maturities in the spread between minimum and maximum yield. Furthermore, the autocorrelation coefficients reveal that yields with longer maturities are more persistent than yields with short maturities, while the Jarque-Bera statistics show that normality is rejected for all maturities. Cross-correlations are very high, ranging between 0.806 and 0.995, and decrease when the difference in maturities becomes larger. A complete overview of the cross-correlations is presented in appendix E. Summarizing our observations for the yields, the data shows convincingly that the yield series exhibit the well known stylized facts of interest rates. The autocorrelations of the level, slope and curvature proxies reveal high persistence in the proxies although persistence is higher for level than for slope and curvature.

3.2 Macroeconomic Data

The macroeconomic dataset consists of monthly observations on 128 macroeconomic time series. The set contains variables from a large variety of sectors of the macroeconomy. For the sake of clarity the variables are sorted into nine economically meaningful categories. Table 2 shows an overview of the categories and the number of variables per category, whereas an complete overview of the macroeconomic dataset is given in appendix A. The dataset is obtained from the Federal Reserve Economic Database and is freely available on the website of the Federal Reserve Bank of St. Louis.[3] If necessary, we transform the series such that they exhibit or approximate stationarity. We do this by either taking (1) first differences, (2) natural log, (3) first difference of natural log, (4) second difference of natural log or (5) first difference of percent change. For the type of transformation per series, we follow the suggestions of McCracken and Ng (2016). Thereafter, we follow Stock and Watson (2002) by considering observations whose absolute median deviation is larger than six times the inter quartile range as outliers, and replacing them by the median value of the last five observations. Finally, we standardize the series by subtracting the mean and dividing by the standard deviation for each series.

Table 2: Categorization of macroeconomic dataset

Category	Number of variables
Output and income	16
Labor market	31
Housing	10
Consumption, orders and inventories	10
Money and credit	14
Interest rates and spreads	17
Exchange rates	5
Prices	20
Stock market	5
Total	128

4 In-Sample Results

4.1 DNS: Dynamic Nelson-Siegel Model

Table 3a shows the estimation results for the autoregressive structure of the Nelson-Siegel factors in the standard DNS model. The diagonal elements in the coefficient matrix are all close to

[3] The set of macroeconomic variables, the code for performing transformations, and a document with the type of transformation per variable are available at https://research.stlouisfed.org/econ/mccracken/fred-databases/

Table 3: AR(1) estimates of the DNS model.

(a) DNS coefficient matrix and constant

	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	μ
$\beta_{1,t}$ (level)	0.984*** (0.0100)	0	0	4.997*** (0.7544)
$\beta_{2,t}$ (slope)	0	0.980*** (0.0105)	0	−2.655*** (0.8097)
$\beta_{3,t}$ (curvature)	0	0	0.945*** (0.0186)	−1.863** (0.7732)

(b) DNS variance matrix

	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$
$\beta_{1,t}$ (level)	0.056*** (0.0060)	0	0
$\beta_{2,t}$ (slope)	0	0.103*** (0.0096)	0
$\beta_{3,t}$ (curvature)	0	0	0.571*** (0.0527)

NOTE: This table reports the estimates of the autoregressive model for the latent factors. Panel (a) shows the estimates for the state coefficient matrix Φ and the constant μ. Panel (b) shows estimates of the state variance matrix Σ_η. T-values are shown between brackets below the estimates. One asterisk (*) indicates a 90% significance level, two asterisks (**) indicate a 95% significance level and three asterisks (***) indicate a 99% significance level.

one. Hence, the high persistence observed in table 1 is confirmed by the estimation results. Observing the variance of the latent factors, we can conclude that the curvature factor contains the highest level of variation and the level factor the lowest. The estimation result of the loading parameter equals $\lambda = 0.046$. This corresponds to a curvature loading attaining its maximum at a maturity of 36 months. The constant for the level factor is close to the average 360-month yield, which we used as proxy for the level factor. Furthermore, the constant for the slope factor is negative, which corresponds to a upward slope. All estimates differ significantly from zero at a 99% confidence level, except the constant for the curvature factor which is significant at a 95% confidence level. Figure 4 displays the DNS factors together with the proxies for the corresponding factors. The level factor shows a steadily declining pattern over time. The slope factor alternates periods of high values with periods of low values including two short periods of negative values, meaning that the yield curve is in the vast majority of cases upward sloping but the steepness of the curve varies considerably. The curvature factor has a mean below zero and shows a slowly decreasing pattern between the period from 1994 to 2014, thereafter the curvature factor starts to increase towards the level at the beginning of the sample period.

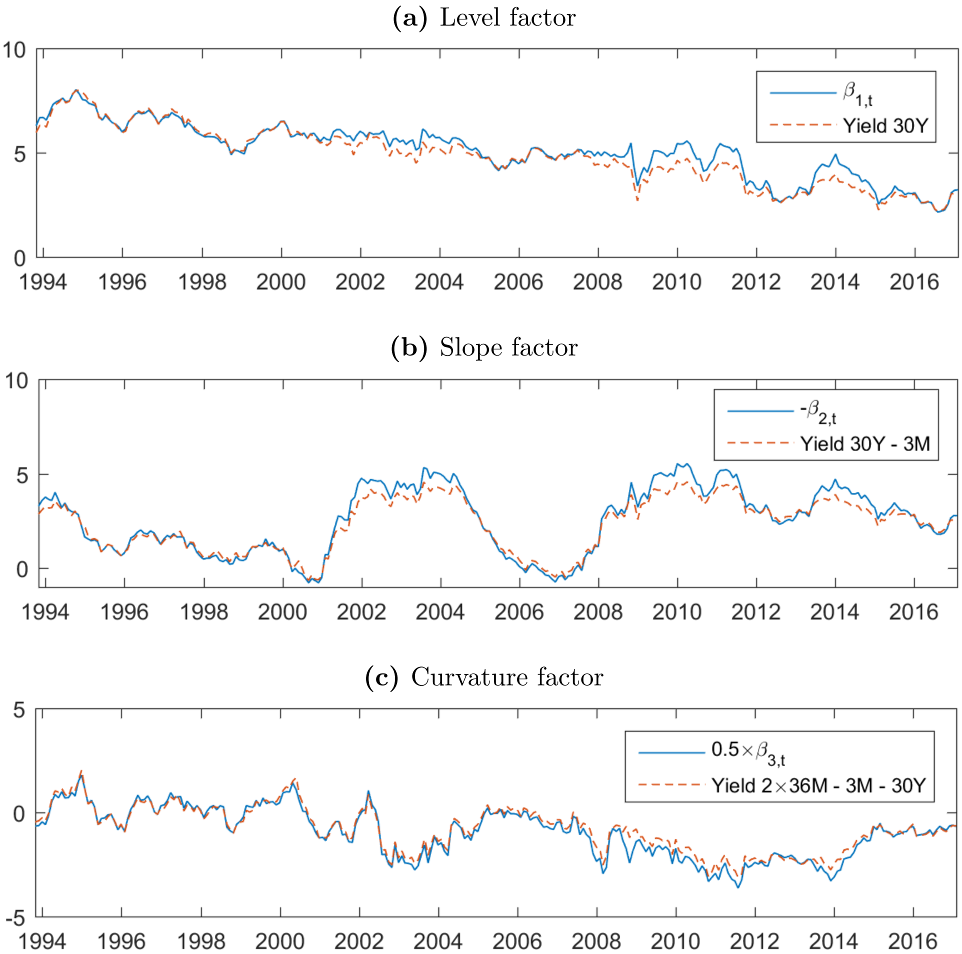


Figure 4: Filtered factors. This figure displays the filtered estimates of the level, slope and curvature factor from the DNS model over the period October 1993 to January 2017. Panels (a), (b) and (c) present the level, slope and curvature factors respectively, together with their empirical proxies. The proxy for level is the 360-month yield, the proxy for slope is the 360-month yield minus the 3-month yield and the proxy for curvature is two times the 36-month yield minus the 3-month and 360-month yield.

4.2 DNS-L: Time-Varying Loading Parameter

Koopman et al. (2010) provide some evidence that the assumption of a constant loading parameter does not necessarily hold by estimating the λ for four equally sized subperiods from the full sample. Applying the same idea to our sample, we obtain four estimates for λ: 0.0663, 0.0490, 0.0483 and 0.0410. The standard errors of these estimates are small enough to conclude that the estimates for all periods (except 2 and 3) differ significantly. Hence, this provides some evidence for the need of a varying loading parameter.

In this section we consider the DNS-L model where the natural logarithm of the loading parameter λ is modeled as a latent factor such as the factors in the autoregressive specification of the DNS model described in section 2.2. We estimate the coefficients and obtain filtered estimates for the factors using the extended Kalman filter as described in section 2.8. The estimated factors show a similar pattern as the factors in the DNS model but differ somewhat

Table 4: AR(1) estimates for the DNS-L model.

(a) DNS-L coefficient matrix and constant

	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	$\log(\lambda_{t-1})$	μ
$\beta_{1,t}$ (level)	0.988*** (0.0080)	0	0	0	4.941*** (0.7856)
$\beta_{2,t}$ (slope)	0	0.983*** (0.0092)	0	0	−2.660*** (0.8050)
$\beta_{3,t}$ (curvature)	0	0	0.953*** (0.0169)	0	−1.586** (0.8035)
$\log(\lambda_t)$	0	0	0	0.844*** (0.0238)	−3.230*** (0.0607)

(b) DNS-L variance matrix

	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$	$\log(\lambda_t)$
$\beta_{1,t}$ (level)	0.036*** (0.0038)	0	0	0
$\beta_{2,t}$ (slope)		0.074*** (0.0070)	0	0
$\beta_{3,t}$ (curvature)			0.422*** (0.0472)	0
$\log(\lambda_t)$				0.019*** (0.0028)

NOTE: This table reports the estimates of the autoregressive model for the latent factors. Panel (a) shows the estimates for the state coefficient matrix Φ^L and the constant μ. Panel (b) shows estimates of the state variance matrix Σ_η^L. T-values are shown between brackets below the estimates. One asterisk (*) indicates a 90% significance level, two asterisks (**) indicate a 95% significance level and three asterisks (***) indicate a 99% significance level.

in certain periods. Differences are most pronounced in the period from 2009 to 2017.[4] The estimated coefficients and variance matrix are presented in table 4. The autoregressive coefficient of the natural logarithm of the loading parameter shows a fairly high level of persistence. All other estimates are very similar to those of the standard DNS model.

The slope- and curvature factor loadings are heavily depending on the loading parameter. Therefore, we take a closer look on the effect of the introduction of a time-varying loading parameter compared to the estimated constant loading parameter of the standard DNS model. As mentioned in section 4.1, the estimated loading parameter for the DNS model is equal to 0.046. Figure 5 shows how the loading parameter evolves in the DNS-L model. It can be noticed that the loading parameter varies substantially over the sample period, moving between values of 0.025 and 0.114. Furthermore, the figure shows a reasonably high level of persistence in the loading parameter which is confirmed by the autoregressive coefficient of 0.844 displayed in table 4a. In figure 6 we plot the slope- and curvature loadings for the minimum and maximum loading parameter achieved in the DNS-L model, together with the loadings corresponding to $\lambda = 0.046$ which was estimated using the standard DNS model. From this figure we can conclude that the factor loadings differ substantially over time. Furthermore, the figure confirms the fact that the exponential decay is stronger and the curvature loading reaches it's maximum at a shorter maturity for a higher λ.

[4]Differences in the factors of different models are generally not of substantial size. For the sake of brevity we therefore do not include a plot of the factors of all models in our analysis.

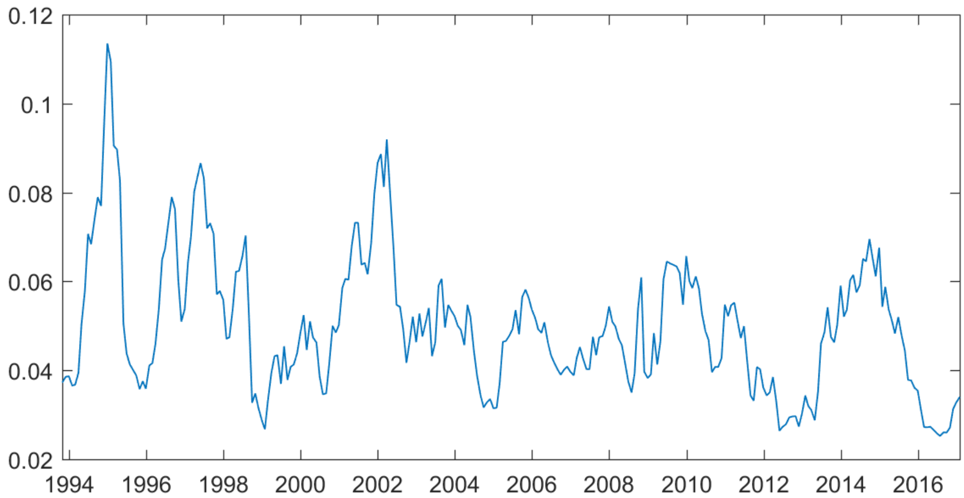


Figure 5: Time-varying loading parameter. This figure presents the estimates of the loading parameter over the period October 1993 to January 2017. The estimates are obtained from the DNS-L model where the loading parameter is treated as a latent factor.

Looking at the yield curves for the periods where the loading parameters achieves a high value, we observe that the yield curves all have the property to be very steep at the short end of the curve (e.g. 1995, 2002). Intuitively a high loading parameter makes perfectly sense in this case since a high loading parameter corresponds to a curvature loading reaching it's maximum at a short maturity (see figure 6b). On the other hand, periods with a small loading parameter correspond with a more flat or gradually increasing curve (e.g. 1999, 2013).

The filtered errors of the DNS-L model are presented in table 5a. Compared to the standard DNS model, the mean absolute error and standard deviation are smaller for seven out of ten maturities. The average mean absolute error and standard deviation across maturities is substantially lower. Improvements in fit are well diversified across the curve. However, the improvement in the fit of the 360-month yield stands out as it shows substantial larger improvement than the other maturities. Table 6 reports the log-likelihood, the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the likelihood ratio statistic for the DNS-L model. The log-likelihood of this model improved by 388 compared to the standard DNS model. This comes roughly down to a impressive 20% improvement. Furthermore, the information criteria and the likelihood ratio statistic confirm the large improvement of fit due to the varying loading parameter. Hence, there is sufficient evidence to conclude that DNS-L model outperforms the standard DNS model in terms of in-sample fit.

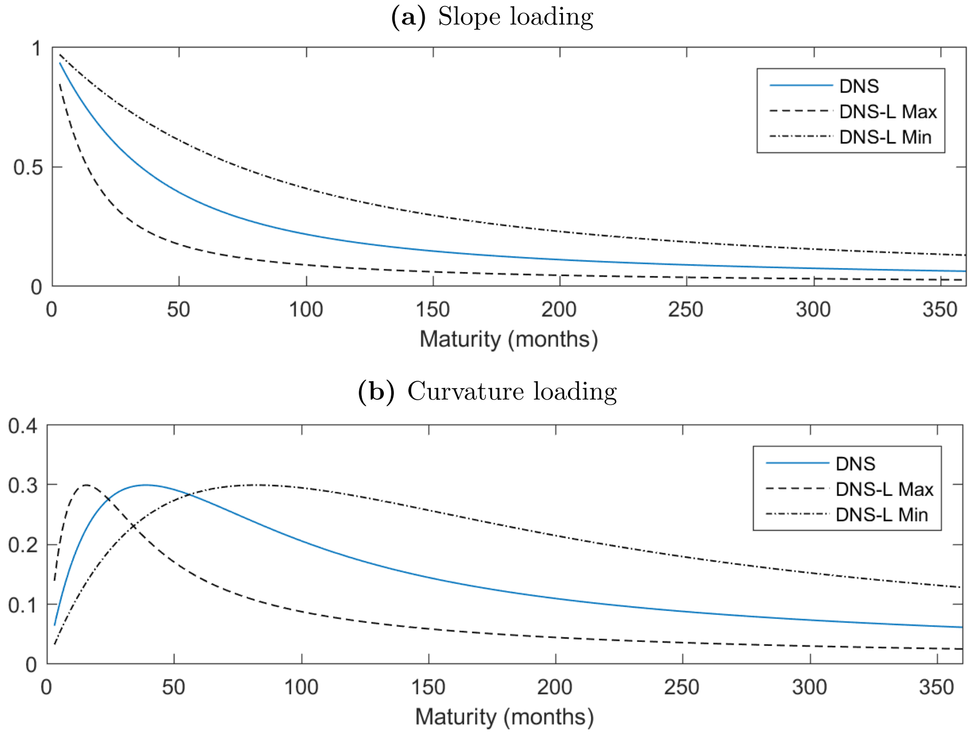


Figure 6: Minimum and maximum loadings of slope and curvature. This figure displays the factor loadings per maturity corresponding to the minimum and maximum loading parameter estimated in the DNS-L model. In addition it shows the factor loadings corresponding with the loading parameter estimated using the DNS model. Panel (a) presents the loadings for the slope factor and panel (b) presents the loadings for the curvature factor.

Table 5: Filtered errors.

(a) Filtered errors of DNS, DNS-L and DNS-V model

	DNS		DNS-L		DNS-V	
Maturity	MAE	Std	MAE	Std	MAE	Std
3	11.28	12.62	10.19	10.91	9.01	11.55
6	0.00	0.00	0.64	1.19	0.16	0.49
12	5.85	7.59	4.27	5.81	4.40	6.11
24	5.26	5.47	3.97	4.04	0.19	0.49
36	0.03	0.04	1.99	2.68	2.56	3.23
60	5.58	6.68	2.92	3.21	0.02	0.00
84	5.31	6.64	2.80	3.51	3.88	5.38
120	4.15	4.50	5.05	5.77	5.03	5.83
240	17.30	15.10	16.31	14.21	14.27	14.11
360	12.77	15.71	5.41	6.81	3.80	4.47
Mean	6.75	7.44	5.35	5.81	4.33	5.17
Median	5.45	6.66	4.12	4.91	3.84	4.93
# Improved			7	7	7	7

(b) Filtered errors of DNS-M(SE), DNS-M(PCA) and DNS-M(PLS) model

	DNS-M(SE)		DNS-M(PCA)		DNS-M(PLS)	
Maturity	MAE	Std	MAE	Std	MAE	Std
3	11.26	12.55	11.28	12.62	11.28	12.62
6	0.61	0.79	0.00	0.00	0.00	0.00
12	5.66	7.32	5.84	7.58	5.84	7.58
24	5.17	5.37	5.26	5.47	5.26	5.47
36	0.00	0.00	0.02	0.03	0.02	0.02
60	5.33	6.43	5.59	6.69	5.59	6.70
84	5.09	6.30	5.29	6.66	5.30	6.67
120	3.30	3.58	4.30	4.70	4.32	4.72
240	17.97	15.58	17.19	15.07	17.17	15.03
360	13.52	16.51	12.81	15.77	12.80	15.75
Mean	6.79	7.44	6.76	7.46	6.76	7.46
Median	5.25	6.37	5.44	6.67	5.45	6.68
# Improved	7	7	6	6	5	5

(c) Filtered errors of DNS-LVM(SE), DNS-LVM(PCA) and DNS-LVM(PLS) model

	DNS-LVM(ES)		DNS-LVM(PCA)		DNS-LVM(PLS)	
Maturity	MAE	Std	MAE	Std	MAE	Std
3	10.22	10.32	10.69	10.63	9.24	10.32
6	0.54	1.42	0.72	1.26	1.65	2.33
12	4.12	5.05	4.37	5.36	3.63	4.99
24	4.30	4.14	4.17	4.12	1.34	1.81
36	0.43	0.99	0.56	0.93	2.35	3.18
60	1.97	2.16	1.37	1.85	0.95	1.44
84	3.13	3.83	3.02	3.69	2.89	3.63
120	3.84	4.82	4.18	5.08	4.62	5.83
240	15.72	13.60	15.16	13.34	12.53	12.24
360	1.62	2.79	2.52	3.33	3.38	4.56
Mean	4.59	4.91	4.68	4.96	4.26	5.03
Median	3.49	3.99	3.60	3.91	3.13	4.10
# Improved	8	7	7	7	7	7

NOTE: This table reports the mean absolute errors (MAE) and standard deviations (Std) of the filtered errors (in basis points) for each maturity. The filtered errors are defined as the observed yields minus the filtered estimates obtained using the (extended) Kalman filter. The table also reports the average and median of the mean absolute errors and standard deviation across maturities as well as the number of maturities for which a model improved on the DNS model. Panel (a) presents results for the DNS, DNS-L and DNS-V model, panel (b) presents results for the DNS-M(SE), DNS-M(PCA) and DNS-M(PLS) model, and panel (c) presents results for the DNS-LVM(ES), DNS-LVM(PCA) and DNS-LVM(PLS) model.

Table 6: Log-likelihood values and information criteria

Model	Log-likelihood	AIC	BIC	# par	LR-Statistic
DNS	1726.7	-3413.5	-3340.8	20	
DNS-L	2114.5	-4187.9	-4107.9	22	775.5
DNS-V	2199.6	-4335.3	-4219.0	32	945.8
DNS-M(ES)	1829.6	-3711.1	-3613.0	27	205.7
DNS-M(PCA)	1744.4	-3558.8	-3686.1	35	35.4
DNS-M(PLS)	1747.0	-3424.0	-3296.8	35	40.5
DNS-LVM(ES)	2203.3	-4345.0	-4232.3	31	953.1
DNS-LVM(PCA)	2295.6	-4487.2	-4298.2	52	1137.8
DNS-LVM(PLS)	2361.7	-4339.9	-4223.6	52	1269.9

NOTE: This tables reports the log-likelihood value, the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the number of parameters for each model. In addition it reports values of the likelihood ratio (LR) test for all extended models against the standard DNS model.

4.3 DNS-V: Time-Varying Volatility

A second extension that we consider for the DNS model is the introduction of time-varying volatility as explained in section 2.4. In the standard DNS model, the yields are captured by the three Nelson-Siegel factors. The remaining errors are assumed to be white noise. The inclusion of a common shock component in the DNS-V model provides more flexibility to the model. In the extended model, we included the common shock component in the measurement equation, see (15). Moreover, these shocks are expected to arrive in clusters of high- and low magnitude. For this purpose, the volatility of the common shock is modeled using a GARCH specification. In periods with low volatility, the shock component is small and the measurement error is still close to white noise. However, when the volatility is high this is not the case. Furthermore, the common shock component has different impact in terms of magnitude on the variety of maturities. Therefore, the vector Γ determines the sensitivity of the yields for maturity i to the common shock component.

The autoregressive coefficients are very similar to those of the standard DNS model and can be found in table 12 in appendix E. The loading parameter is estimated at 0.0513 with a standard error equal to 0.0009. Comparing the loading parameters (and the standard errors) of both models it becomes clear that the optimal loading parameter is affected by the introduction of the common shock component. Thus, the volatility component affects the estimation also indirectly through λ. Since the optimal loading parameter is higher in the DNS-V model, it means that the exponential decay is stronger and the maximum of the curvature factor is reached at a shorter maturity.

Table 7 displays the estimated GARCH parameters and the corresponding standard errors. Since we estimate all loading components in Γ it is not possible to identify γ_0, therefore it is fixed at a small value. For the remaining GARCH parameters γ_1 and γ_2 it holds that γ_1 is significant at a 95% confidence level and γ_2 at a 99% confidence level. The moderate value and the significance of the parameters indicate that the volatility of the common shock component is determined by the innovations as well as the previous volatility estimate. Hence, the volatility level changes gradually over time and therefore the model captures the feature of volatility clustering.

Table 7: GARCH parameters

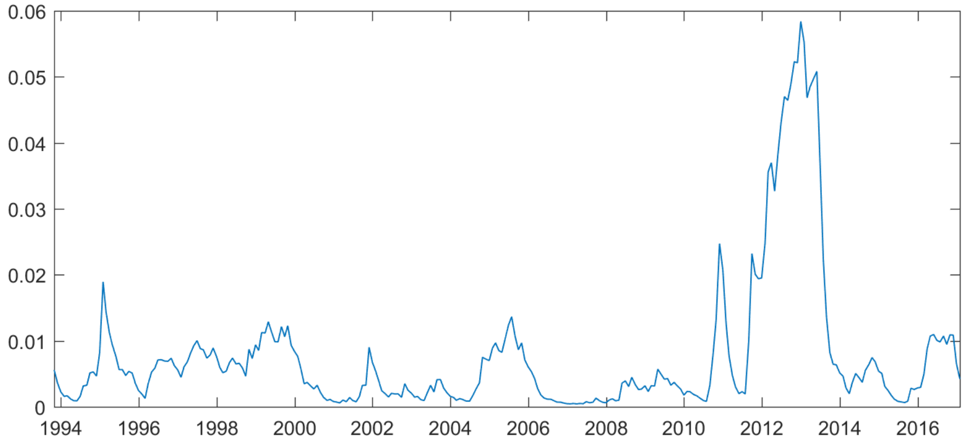
γ_0	γ_1	γ_2
0.0001	0.406**	0.576***
(*fixed*)	(0.1795)	(0.1763)

NOTE: This table reports the estimated parameters of the GARCH specification in the DNS-V model. One asterisk (*) indicates a 90% significance level, two asterisks (**) indicate a 95% significance level and three asterisks (***) indicate a 99% significance level.

Figure 7a plots the volatility of the common shock component over time. It shows that the volatility is particularly high in the period between 2011 and 2014. During this period the Federal Open Market Committee was performing active policy on the yield curve by launching a Quantitative Easing program. Speculations about the length and size of the program resulted in uncertainty and a higher volatility in the bond market. Other (smaller) peaks can also be related to events with large economic impact, such as the Mexican Peso crisis (1994), the collapse of the dot-com bubble (1999) and brexit (2016). The elements of Γ are presented graphically in figure 7b. Remarkable is that the maturities 12- and 24-month yields seem to be insensitive for the common GARCH component, while the medium term maturities of 60, 84 and 120 months are the most sensitive to the GARCH component.

Table 5a reports the mean absolute error and standard deviation of the filtered errors per maturity. Compared to the standard DNS model, the DNS-V model improves on seven out of ten maturities in terms of the mean error as well as in terms of standard deviation. Furthermore, the average mean absolute error and standard deviation across the different maturities has greatly improved. This is mostly due to improvements in the fit of the 60- and 360-month yields. The 60-month yield is now almost perfectly fitted, while the 6- and 24-month yields are also close to a perfect fit in this model. However, for the 36-month yield the fit is clearly worse than in the standard DNS model. Next, we compare the log-likelihood and information criteria of the DNS-V model with the standard DNS model in table 6. The difference between the models is equal to 488 in the advantage of the DNS-V model. This improvement is even more convincing than that of the DNS-L model. After correcting for the additional parameters included in the DNS-V model, the information criteria are also better for the DNS-V model. Hence, there is enough evidence to conclude that the extension with varying volatility drastically improves the fit and that the improvement is even more pronounced than for the extension with a time-varying loading parameter.

(a) GARCH volatility component h_t



(b) Volatility loading per maturity

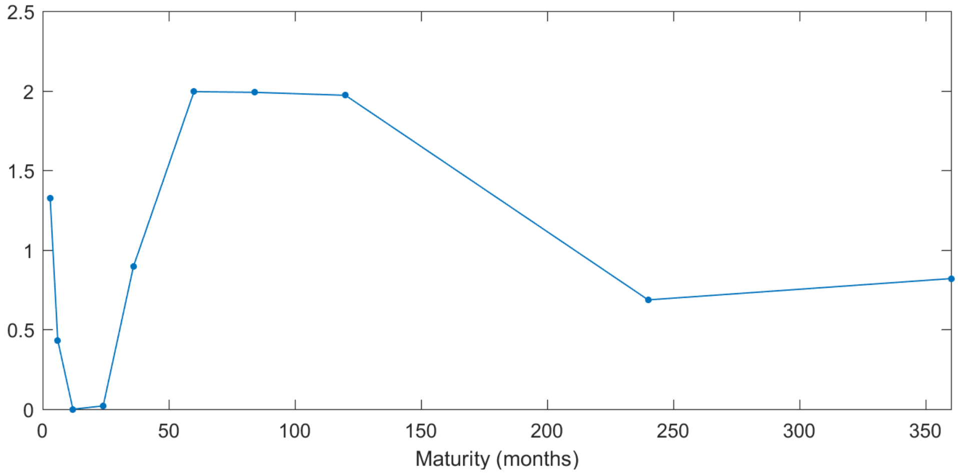


Figure 7: Time-varying volatility and sensitivity. This figure presents the time-varying volatility. Panel (a) plots the volatility h_t of the common shock component ε_t^* from the DNS-V model over the period October 1993 to January 2017. Panel (b) plot the elements of Γ against the maturities. These elements present the sensitivity of a yield to the common shock component ε_t^*.

4.4 DNS-M: Time-Varying Unconditional Mean

The last modification to the standard DNS model that we consider is the allowance of a time-varying unconditional mean. Dijk et al. (2014) argue that given the behavior of the US Treasury yields, a constant unconditional mean may be regarded as inappropriate. To obtain some indication whether this argument holds on the sample we use in this study, we obtain estimates for μ using four subperiods as explained in section 4.2. Estimates for the unconditional mean of the level factor differ convincingly. Differences for the slope and curvature factors are also clearly visible but are somewhat less pronounced.

For the FADNS model we construct macroeconomic factors using PCA and PLS as described in section 2.5.2. The goal of PCA is to capture as much as possible variation of the macroeconomic dataset, while PLS constructs factors to explain as much as possible variation of the response variable. Applying both methods to the full (transformed) set of macroeconomic variables with 128 variables yields rather disappointing results, i.e. the first three principal components explain only 34% of the variation in the dataset and the first three factors con-

(a) Output & Income **(b)** IR & ER **(c)** Prices

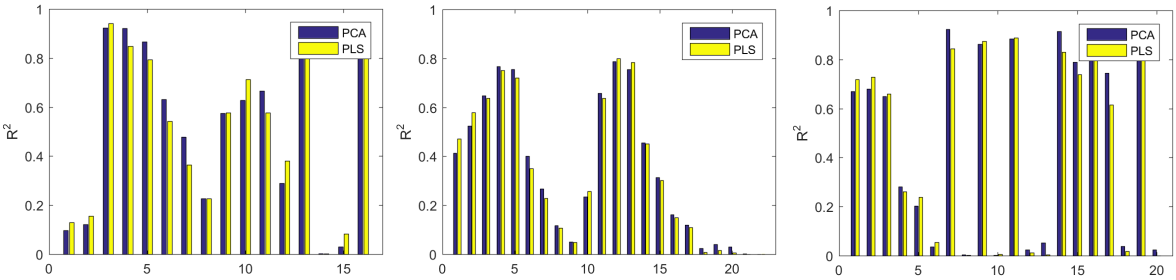
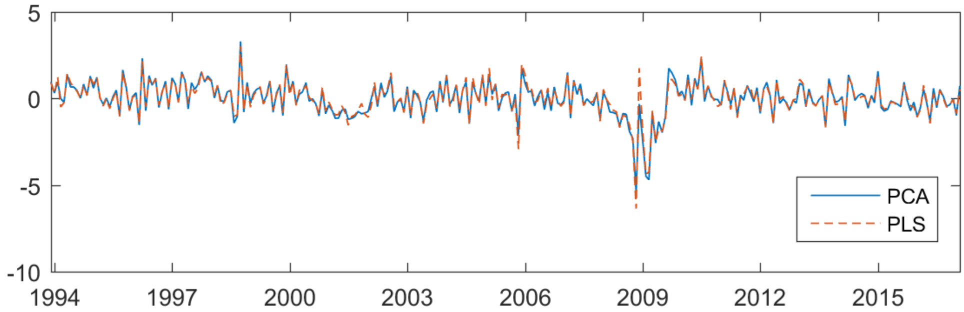


Figure 8: R^2-values in regressions of individual macro series on common macro factors. This figures show the r-squared values obtained from regressions of individuals variables in a certain group on the first common component of this group, obtained by either PCA or PLS. Panels (a), (b) and (c) show the results for the groups "Output and income", "Interest rates and exchange rates" and "Prices" of the complete macroeconomic dataset, respectively. The groups "Output and income", "Interest rates and exchange rates" and "Prices" contain 16, 22 and 20 variables, respectively. R-squared values using a common factor obtained by PCA are presented using blue bars, the yellow bars represent the same values for a factor obtained using PLS.

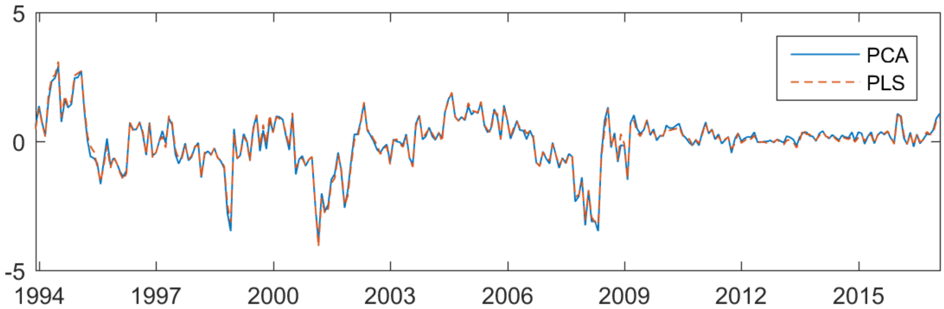
structed by PLS explain only 27%. In other words, the factors contain a rather small part of the information available in the dataset. A possible reason for this result may be that the variables in the dataset are too diverse, such that it is hard to summarize the information into a small number of factors. As it is not desirable that the factors contain such little information we choose to apply PCA and PLS in a different approach, i.e. we apply PCA and PLS to three subsets of the macroeconomic dataset that hold the most important information and contain somewhat similar variables. Then, we use the first common factor of these groups in our analysis. Doing this, we expect PCA and PLS to capture a substantial higher percentage of variation. The subsets where we apply PCA and PLS to are the groups 'Output and income' (group one), 'Interest rates and exchange rates' (group six) and 'Prices' (group seven), following Exterkate et al. (2013). This choice is motivated by the fact that the factors production, inflation and the federal funds rate are often considered as the most influential variables to the yield curve. The considered groups all contain variables that are somewhat related to these factors and thereby are expected to contain the most important information.

For groups one, six and seven the first common factors explain 51%, 39% and 48%, respectively, while the the common factors using PLS explain respectively 50%, 39% and 46%. Figure 8 shows the R^2 when regressing individual series of a group on the common factor of this group. Normally, when applying factor extraction methods to a large set of variables the r-squared values help to give interpretation to the constructed factor. Here, we selected the groups to which we apply the factor construction methods up front. Hence, the interpretation of the factor is already clear. However, it is interesting to see which variables contribute the most to the constructed factor. In group one we observe large value for the indices: total, products, final products, manufacturing and capacity utilization, suggesting that mainly the large indices are important. For group six, the figure shows particularly high values for short term treasury rates and their spreads with the federal fund rate. On the other hand, values for exchange

(a) First common factor "Output and income"



(b) First common factor "Interest rates and exchange rates"



(c) First common factor "Prices"

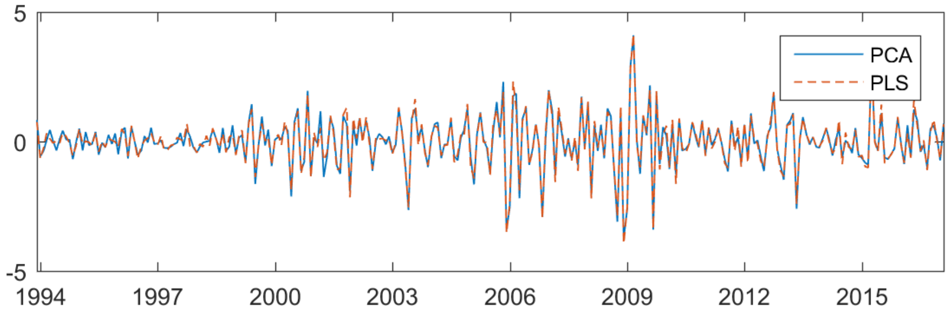


Figure 9: Time series plots of constructed common macroeconomic factors. This figure shows the time series plots of the first common factors, constructed using either PCA or PLS. Panels (a), (b) and (c) plot the first common factor for the groups "Output and income", "Interest rates and exchange rates" and "Prices" from the complete transformed macroeconomic dataset, respectively.

rates are close to zero, meaning that the factor of group six contains almost no information of these variables. Finally, we observe that the common factor of group seven mainly resembles price indices capturing (almost) all goods and price indices for transportation and commodities. Furthermore, we observe small differences between the r-squared values using PCA and PLS, meaning that the contribution of the variables differs for the different factor construction methods. Figure 9 shows the time series plots of the first common factors, obtained by applying either PCA or PLS to the earlier discussed groups. Remarkable is that the factors obtained using PCA are very similar to those obtained using PLS. However, differences are notable.

For the DNS-M(ES) model the estimation result of the smoothing parameter α equals 0.1964. This value is substantially higher than the fixed value adopted by Dijk et al. (2014) and results in a mean that is less smoothed and more sensitive to new factor observations.

Table 5b presents the mean absolute error and standard deviations of the filtered errors per maturity. Comparing the different variants of the DNS model with a time-varying unconditional mean, we observe that the errors of these model are very similar. Especially, the results of the DNS-M(PCA) and DNS-M(PLS) models are almost identical. For the DNS-M(ES) model we observe a slightly higher bias for the longer maturities. The errors of all variants are very similar those of the standard DNS model as well. Hence, the error analysis does not support an improvement in the in-sample fit when using a time-varying unconditional mean. However, table 6 shows improvements in the likelihood compared to the likelihood value of the standard DNS model, although the improvements are much smaller than those of the DNS-L and DNS-V model. In particular, the improvements of the DNS-M(PCA) and DNS-M(PLS) are small. However, these improvements are still significant.

4.5 DNS-LVM: Combination of Extensions

In the subsections above we provided a extensive in-sample analysis on all models with one specific extension for the DNS model. Therefore, we only address the performance of the multi-extension models based on the filtered errors and the likelihood values and information criteria.

As the results of the previous subsections were fairly encouraging we next discuss the in-sample performance of the models where we combine extensions in order to find out whether the combination of extension can further improve the in-sample fit. Table 5c presents the filtered errors of all variants of the DNS-LVM model. Comparing the filtered errors of the DNS-LVM models with those of DNS model we observe that the mean absolute error as well as the standard deviations of the DNS-LVM(PCA) and DNS-LVM(PLS) models are smaller for seven out of ten maturities. The number of improvements is equal to that of the DNS-L and DNS-V model. However, the average standard deviation for both models has improved compared to that of the DNS-L and DNS-V models. Comparing the DNS-LVM(PLS) model with the DNS model, we observe the filtered mean absolute error is lower for eight out of ten maturities, whereas the standard deviation of the filtered errors is smaller for seven out of ten maturities. Such improvement in the mean filtered errors has not been achieved by any of the other models.

The log-likelihood values and information criteria of the DNS-LVM models are reported in table 6. The table shows significant improvements for all variants of the DNS-LVM model compared to the standard DNS model. However, we are particularly interested whether the multi-extension models perform better than the best single-extension model. Therefore, we compute the likelihood-ratio statistic for each of the DNS-LVM models against the DNS-V model. These values equal 7.3, 191.9 and 324.1 for the DNS-LVM(ES), the DNS-LVM(PCA) and DNS-LVM(PLS) model, respectively. All values indicate significant improvements. Considering all results for the DNS-LVM models we conclude that combining individual extensions contributes to significant further improvement in fit of the DNS model.

5 Out-of-Sample Forecasting

In this section we examine the forecasting ability of all models presented in section 2 for a 1-, 3-, 6- and 12-month forecast horizon over different (sub)samples. For many of these models, only an in-sample analysis is available, e.g. Koopman et al. (2010) only present an in-sample analysis of the DNS-L and DNS-V model. However, besides describing and fitting the yield curve, the DNS model may be particular interesting for forecasting the yield curve. In this section we aim to present a extensive overview of the forecasting ability of the DNS model and all its discussed extensions.

5.1 Description

We divide our dataset in an initial estimation sample that covers the period from 1993/10 to 2003/9 (120 observations) and a forecasting sample that covers the period from 2003/10 to 2017/1 (160 observations). Moreover, we divide the forecasting sample into two subsamples covering the periods 2003/10 to 2010/6 (80 observations) and 2010/7 to 2017/1 (80 observations), such that we obtain a total of three subsamples. To construct forecasts, we estimate the models recursively using a moving window of 120 observations. More specifically, we start the procedure by estimating the parameters of the different models over the first subsample and use the parameter estimates and the prediction step in the Kalman filter to construct forecasts for the different horizons. Next, we exclude the first observation of the estimation window and include the upcoming out-of-sample observation. Then, we re-estimate the parameters and construct a new set of forecasts. To construct the forecast, we first obtain the filtered latent factors from the Kalman filter. Then, we predict the h-months ahead state forecast by iteratively applying the prediction step h times to the last filtered observation of the filtered factor. More formally, using the general state space representation, we construct the h-month ahead forecast as

$$\hat{x}_{t+h|t} = \sum_{i=0}^{h-1} A^i C + A^h \hat{x}_{t|t}, \tag{41}$$

where h is the forecast horizon. The optimal h-period ahead point forecast is equal to the following conditional expectation

$$\hat{y}_{t+h} = \mathrm{E}[y_{t+h}|\mathcal{I}_t] = H\hat{x}_{t+h|t}. \tag{42}$$

Hence, we construct the h-period ahead forecast by substituting the predicted state vector from (41) into the measurement equation.

5.2 Evaluation

One of the aims of this study is to discover whether the extensions to the standard DNS model improve the forecast accuracy. In addition, we are interested in the difference in performance between different approaches for a unconditional mean and in the effect of combining extensions to the DNS model. For this purpose we use the root mean squared prediction error (RMSPE) as measure of forecast accuracy. Let $\hat{y}_{t+h|t}(M,\tau)$ denote the h-period ahead point forecast of model M for maturity τ. Then, the RMSPE for time t is given by

$$RMSPE = \sqrt{\frac{1}{S}\sum_t \left(\hat{y}_{t+h|t}(M,\tau) - y_{t+h}(\tau)\right)^2}, \tag{43}$$

where S is the number of predictions that is considered. The RMSPE measures the difference between the observations and the forecasts and should therefore be as small as possible. To test statistical significant outperformance of a model relative to the DNS model, we follow Exterkate et al. (2013) and many others by using the Diebold-Mariano test proposed by Diebold and Mariano (1995).[5]

Let $\hat{y}^1_{t+h}$ and $\hat{y}^2_{t+h}$ be two competing h-period ahead forecasts for y_{t+h}. The forecast errors of both models are then given by

$$\varepsilon^1_{t+h} = y_{t+h} - \hat{y}^1_{t+h}, \qquad \varepsilon^2_{t+h} = y_{t+h} - \hat{y}^2_{t+h}.$$

To measure the forecast accuracy of model i a particular loss function is used, i.e.

$$g\left(y_{t+h}, \hat{y}^i_{t+h}\right) = g\left(\varepsilon^i_{t+h}\right), \quad \text{for } i = 1, 2.$$

Here, we choose to use the very popular squared error loss function. To determine which models performs better, consider the loss-differential series

$$d_t = g\left(\varepsilon^1_{t+h|t}\right) - g\left(\varepsilon^2_{t+h|t}\right).$$

If a model performs better than the other, the loss differential should differ significantly from zero. To test whether model two outperforms model one, we take as null hypothesis $H_0 : \mathrm{E}[d_t] = 0$. The Diebold-Mariano test statistic is

$$D = \frac{\bar{d}}{\sqrt{\frac{V(\bar{d})}{T}}}, \tag{44}$$

where $\bar{d} = \frac{1}{S}\sum_t d_t$ and $V(\bar{d})$ is a consistent estimate of the long run asymptotic variance of $\bar{d}$ and is computated as $V(\bar{d}) = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j$, with $\gamma_j = cov(d_t, d_{t-j})$. Finally, Diebold and Mariano (1995) show that under the null hypostudy $D \overset{A}{\sim} \mathcal{N}(0,1)$, such that a one- or two-sided test can simply be performed.[6] Here, we choose to perform one sided test of all extended models against the standard DNS model. Hence, the DNS model acts as benchmark such that we are

[5] Normally, the Diebold-Mariano test should not be used for comparing the predictive performance of nested models. However, Giacomini and White (2006) prove that the test is valid when using a rolling estimation window as we do in our analysis.

[6] A special remark should be made about the value of $V(\bar{d})$, which can achieve negative values in rare cases. In these cases we set $V(\bar{d})$ to a very low value such that null hypothesis is automatically rejected.

better able to draw conclusions about the gain in forecasting performance of extensions in the standard DNS model.

5.3 Forecast results

Table 8 reports forecasting results of all considered models for the full out-of-sample period from October 2003 to January 2017, whereas table 9 and 10 report results for the subperiods October 2003 to May 2010 and June 2010 to January 2017, respectively. For each table, panels A, B, C and D report results for respectively 1-, 3-, 6- and 12-month horizons. The first row of each panel contains absolute RMSPE values for the DNS model, whereas the remaining rows contain RMSPE values relative to the DNS model. Relative values that are below one and thus outperform the DNS model are indicated with a shaded cell. Bold values indicate significant outperformance of a particular model compared to the DNS model at a significance level of 95%.

Consistent with Exterkate et al. (2013) and others, the tables show the general pattern that forecasts are more accurate when the forecast horizon is shorter, as can be seen from the first row of each panel. This makes intuitively sense since the yields are subject to more information for a longer horizon and thereby are more difficult to predict.

Considering the results for models with one specific extension over the full out-of-sample period (table 8) and at a 1-month horizon, we observe that the DNS-L model is able to outperform the DNS model for all maturities and the DNS-M(ES) model performs better for the majority of maturities. Remarkable however, is that the other models that allow for a time-varying mean (using macroeconomic information) perform worse than the DNS model for all maturities. This may be explained by the fact that these variants of models using a time-varying unconditional mean include substantially more parameters making it harder to correctly estimate the parameters. The improvements of the DNS-L model are significant for all maturities except 120 and 240 months, whereas they are for the DNS-M(ES) model significant for five out of eight improvements. The DNS-V model does significantly improve for the shortest maturity but performs substantially worse for other maturities. This may be explained by the fact that the yields are most volatile at the short end of the curve and time-varying volatility may therefore be only beneficial at this part of the curve. The inclusion of macroeconomic information does not seem to improve forecasts at the shortest horizon as results are slightly worse for all maturities excepts three months. The best performing model at the 1-month horizon is clearly the DNS-L model.

For the 3-month horizon the results are pretty similar to the results at a 1-month horizon. The DNS-L and the DNS-M(ES) models show again significant improvements for most maturities. In fact the DNS-M(ES) model does now even improve on the DNS model for all maturities. Furthermore, the DNS-V model significantly improves for the three shortest maturities and does perform for the longer maturities in relative terms less worse then at a 1-month horizon. The performance of the DNS-V model is at this horizon even slightly better than the DNS-L model.

For the 6-month horizon forecasts, the models using macroeconomic information start to outperform the DNS model, suggesting that macroeconomic information is particularly useful for predictions with longer horizons. Although the DNS-M(PLS) model improves just at two maturities, the results for the DNS-M(PCA) model are more promising with improvements on seven out of ten maturities. However, all improvements from the models using macroeconomic information are small and not significant. The DNS-V model improves at this horizon also for the 24-month maturity. For this horizon it is now quite convincingly the best performing model for short maturities, although it remains substantially worse than the benchmark model for maturities longer than 24 months. For these maturities, the DNS-L model remains convincingly the best performing model. In addition, we observe that the outperformances of the DNS-M(ES) are for the 6-month horizon no longer significant.

The results for the longest horizon are the most promising as can be seen from the high number of highlighted cells. The DNS-L, DNS-M(ES) and DNS-M(PCA) model now all show improvements for all maturities, although improvements of the DNS-M(ES) model are only significant for the longest three maturities and those of the DNS-M(PCA) are all not significant. The DNS-V model does now outperform the benchmark model for five out of ten maturities but remains performing worse for longer maturities. Furthermore, the DNS-M(PLS) model clearly performs worse than the DNS-M(PCA) model as it is still worse than the DNS model for seven out of ten maturities.

Considering the forecast results for the models where we combine all types of extensions described in section 2 into one model, we find fairly promising results. For the 1-month horizon the models are superior to all single-extension models for the 3- and 6-month maturity. The DNS-LVM(ES) model is the overall best performing model for maturities up to 24 months, showing large significant improvements. However, for longer maturities the results are less encouraging as the models using macroeconomic information do not improve the DNS model and the DNS-LVM(ES) model does perform worse than the DNS-L model for the medium term maturities. This is likely due to the bad performance of the time-varying volatility extension for the longer maturities. At the 3-month horizon all the models are superior to the models with individual extensions up to the 24-month maturity and improve the DNS up to the 36-month maturity, whereas the DNS-LVM(ES) model improves the DNS model for all maturities. However, the DNS-L model remains superior for maturities longer than 60-months. The most encouraging results are found for the 6- and 12-month horizons as all models with combined extensions outperform the DNS model for all maturities. Moreover, all three models perform for the majority of maturities better than all models with one specific extensions in terms of RMSPE. For the 6-month horizon improvements vary up to 21% while for the 12-month horizons improvements achieve values up to 34%. However, the combined models achieve less significant values than the DNS-L model.

Summarizing the results for the period October 2003 to January 2017, we find that the DNS-L is overall the best performing single-extension model by showing significant improvements for almost all maturities at all horizons. The DNS-V model seems to be the most appropriate model for forecasting the short- and volatile end of the yield curve. The allowance for a time-varying unconditional mean seems beneficial for all horizons as the DNS-M(ES) model shows improvements, from which many significant, for almost all maturities and all horizons. However, a time-varying mean using macroeconomic information seems substantially less beneficial. Including macroeconomic information using PCA results in slightly better performance than when using PLS. Combining extensions in one model is very beneficial at the short end of the yield curve at forecast horizons of one and three months, whereas it is very beneficial for the complete curve at horizons longer than three months. For the multi-extension models using macroeconomic information, the PLS variant seems to be the better choice in terms of RMSPE and the number of significant improvements, which is remarkable since the PLS variant of the individual extensions performed slightly worse then the variant using PCA. The overall best choice is probably the DNS-LVM(ES) model judging on its impressive relative RMSPE values and high number of significant improvements. However, the more parsimonious DNS-L model may also be a good choice as it achieves significant improvements for almost all maturities at all forecast horizons.

Although table 8 shows that many extended models are able to outperform the standard DNS model for various maturities and forecast horizons, it does not need to hold that these models outperform the benchmark model for the entire sample period. In fact it would be possible that a large part of the improvement in forecast performance comes from a small part of the sample. To examine the robustness of the above discussed results we consider tables 9 and 10 which present the (relative) RMSPE results for the subsamples October 2003 to May 2010 and June 2010 to January 2017. By reconsidering figure 3, we can regard these periods as a rather volatile and stable period. Hence, by observing RMSPE results for these subperiods we are able to get a better insight under which circumstances particular models perform well.

The first thing that stands out when comparing tables 9 and 10 is that the RMSPE values are in general substantially larger for the first, more volatile period. However, the difference is bigger for the shorter maturities than for the longer maturities. This can be explained by the fact that the difference in volatility between the two periods is larger for the short-end of the yield curve than for the long-end of the curve as can be seen from figure 3b. In fact, after calculating the standard deviations for every maturity of the yields in both subperiods we observe that except for the longest two maturities the yields in the second subperiod are more volatile than in the first subperiod. Furthermore, table 10 shows generally more highlighted cells and bold values, suggesting that the extensions are more effective in this period.

Now let me zoom in on the performance of specific models in both subsamples. The DNS-L model shows for both subperiods improvements for almost all maturities. Hence, we can conclude that the DNS-L model is a robust and interesting extension for the DNS model. However, it should be noticed that the improvements for the volatile period are far less convincing than the impressive values in the more stable period. This implies that the introduction of a time-varying loading parameter is particularly a useful extension in a stable period. The DNS-V model does in both subperiods generally improve on the DNS model for short maturities at horizons longer than three months. However, similar to the DNS-L model, improvements are much stronger for the more stable subperiod. Although forecasts for the long-end of the curve relative to the DNS are in the first subperiod better than for the second subperiod, they are still worse than those of the standard DNS model. Considering the model with a time-varying unconditional mean, we find that the DNS-M(ES) model is a robust model as it generally outperforms the DNS model for the majority of maturities and all horizons in both subperiods. Again, this extension to the DNS model is more beneficial in the relatively calm subperiod. In fact, this model is outperformed by the models including macroeconomic information in the first subperiod for the short and most volatile part of the yield curve. We observe that the models including macroeconomic information perform particularly well for maturities with high volatility. In the first subperiod, these are the short maturities whereas for the second subperiod these are the two longest maturities. This result is consistent with Exterkate et al. (2013) who also find that macroeconomic information is particularly useful when volatility is high. Moreover, we observe that the DNS-M(PLS) model performs better than the DNS-M(PCA) model in this case, whereas it performs worse when volatility is low. As the predictions of the DNS-L, DNS-V and DNS-M(ES) models are all particularly beneficial in the relatively calm subperiod, we expect the DNS-LVM(ES) model to be the best performing combined model in this period if the positive effects of the extensions are complementary. Observing table 10 we find that this is indeed the case. The DNS-LVM(ES) model shows very high improvements up to 75% and it outperforms every single-extension model for all maturities and all horizons, while it outperforms the combined extensions models including macroeconomic information in almost every case. The slightly worse performance of the DNS-LVM(PCA) and the DNS-LVM(PLS) model can be explained by the result of the single-extension models, that macroeconomic information worsens the forecast performance in the calm period. However, both models outperform all single extension models on the shortest half of the curve and outperform the DNS model for all maturities at all forecast horizons except one month. For the more volatile period we expect the multi-extension model with macroeconomic to outperform the DNS-LVM(ES) model for the maturities where macroeconomic information was beneficial for the single-extension models. For the DNS-LVM(PLS) model this is indeed the case, however for the DNS-LVM(PCA) it is not. Finally, in the calm subperiod the multi-extension models generally outperform the single-extension models for the shortest half of the curve, but are not able to outperform the more parsimonious DNS-L model for the longest half of the curve.

Summarizing the results for the two considered subperiods we find that the DNS-L and DNS-M(ES) models are beneficial extensions in the calm subperiod as well as the volatile period for the complete curve, whereas the DNS-V model is only beneficial for short maturities. However, improvements of these models compared to the DNS model are a lot more pronounced for the calm period. Next, we find that the inclusion macroeconomic information is only beneficial when yields contain a high level of volatility. Furthermore, we find that most positive effects of the single-extension models are complementary when combining extensions into one model. This results in very large improvements for the multi-extension models in the calm subperiod and to somewhat smaller improvements in the volatile subperiod. Finally, for the calm period the DNS-LVM(ES) is without doubt the overall best performing model, while for the volatile subperiod the DNS-LVM(PLS) model is best for short maturities and the DNS-L model is best for longer maturities.

Table 8: Root mean squared prediction errors, 2003/10-2017/1

Model / Maturity	3	6	12	24	36	60	84	120	240	360
Panel A: 1-month horizon										
DNS	29.8	20.8	19.1	23.5	26.0	27.0	26.8	25.3	28.1	28.2
DNS-L	**0.95**	**0.96**	**0.92**	**0.93**	**0.95**	**0.95**	**0.96**	0.97	0.98	**0.90**
DNS-V	**0.98**	1.01	1.15	1.55	1.97	2.39	2.49	2.52	1.96	1.98
DNS-M(ES)	**0.96**	0.99	0.99	0.99	**0.96**	**0.96**	**0.97**	**0.97**	1.00	1.01
DNS-M(PCA)	0.99	1.00	1.06	1.03	1.02	1.02	1.02	1.03	1.06	1.03
DNS-M(PLS)	0.99	1.05	1.11	1.05	1.04	1.03	1.04	1.05	1.06	1.08
DNS-LVM(ES)	**0.86**	**0.86**	**0.86**	**0.91**	0.97	1.03	1.08	1.10	**0.92**	0.98
DNS-LVM(PCA)	**0.91**	0.90	1.00	1.08	1.20	1.33	1.39	1.44	1.09	1.28
DNS-LVM(PLS)	**0.90**	0.91	1.00	1.05	1.18	1.30	1.36	1.39	1.11	1.19
Panel B: 3-month horizon										
DNS	53.5	47.2	45.5	49.3	52.4	51.8	50.1	47.2	45.4	46.0
DNS-L	**0.95**	**0.95**	**0.93**	**0.91**	**0.90**	**0.90**	**0.92**	**0.92**	**0.96**	**0.95**
DNS-V	**0.93**	**0.93**	**0.94**	1.09	1.27	1.48	1.55	1.57	1.43	1.42
DNS-M(ES)	**0.95**	0.98	0.98	0.97	0.96	**0.95**	0.95	**0.95**	0.97	0.98
DNS-M(PCA)	1.00	0.99	1.01	1.02	1.01	1.01	1.01	1.01	1.02	1.01
DNS-M(PLS)	1.00	1.02	1.03	1.02	1.02	1.02	1.03	1.03	1.03	1.05
DNS-LVM(ES)	**0.79**	**0.80**	**0.81**	**0.85**	**0.87**	0.91	0.96	0.97	0.93	0.95
DNS-LVM(PCA)	0.88	0.85	0.86	0.89	0.93	1.00	1.06	1.09	0.99	1.06
DNS-LVM(PLS)	**0.85**	**0.84**	**0.84**	0.89	0.94	1.02	1.07	1.08	1.02	1.04
Panel C: 6-month horizon										
DNS	84.9	79.6	76.8	77.7	79.3	76.9	73.1	69.0	63.7	63.4
DNS-L	**0.96**	**0.95**	**0.93**	**0.90**	**0.88**	**0.87**	**0.88**	**0.89**	**0.93**	**0.95**
DNS-V	**0.92**	**0.92**	**0.91**	0.98	1.08	1.21	1.26	1.27	1.21	1.22
DNS-M(ES)	0.96	0.98	0.97	0.96	0.95	0.93	0.93	0.92	0.94	0.94
DNS-M(PCA)	0.96	0.97	0.98	0.99	1.00	1.00	1.00	0.99	0.99	0.99
DNS-M(PLS)	1.00	0.99	0.99	1.00	1.01	1.03	1.04	1.04	1.03	1.04
DNS-LVM(ES)	**0.79**	**0.80**	**0.79**	**0.80**	**0.80**	**0.84**	0.89	0.91	0.91	0.92
DNS-LVM(PCA)	**0.86**	**0.85**	**0.83**	**0.82**	**0.83**	0.88	0.93	0.95	0.92	0.99
DNS-LVM(PLS)	**0.80**	**0.81**	**0.80**	**0.81**	**0.84**	**0.90**	0.95	0.96	0.96	0.99
Panel D: 12-month horizon										
DNS	142.0	135.7	129.9	125.4	122.8	113.8	105.7	98.8	89.7	88.3
DNS-L	0.97	0.97	**0.95**	**0.91**	**0.87**	**0.84**	**0.83**	**0.84**	**0.88**	**0.91**
DNS-V	0.94	**0.93**	**0.91**	**0.94**	0.98	1.06	1.09	1.10	1.08	1.09
DNS-M(ES)	0.96	0.97	0.97	0.96	0.94	0.92	0.91	**0.89**	**0.88**	**0.88**
DNS-M(PCA)	0.95	0.95	0.96	0.97	0.98	0.98	0.99	0.99	0.99	0.99
DNS-M(PLS)	1.00	0.99	0.98	0.99	1.00	1.02	1.04	1.04	1.04	1.05
DNS-LVM(ES)	0.87	0.86	0.82	0.75	**0.69**	**0.66**	**0.70**	**0.74**	**0.76**	**0.77**
DNS-LVM(PCA)	0.98	0.95	0.89	0.82	0.78	0.78	0.81	0.83	0.85	0.92
DNS-LVM(PLS)	0.83	**0.82**	**0.79**	**0.77**	**0.78**	**0.83**	**0.87**	**0.88**	0.91	0.93

NOTE: This table reports the forecast performance of all models described in section 2, for the period 2003/10 to 2017/1. Panels A, B, C and D present results for a 1, 3, 6 and 12 month(s) horizon, respectively. The first row of each panel reports the RMSPE of the DNS model as described in (43) in basis points. The remaining rows report RMSPE values relative to the DNS model. Shaded cells indicate relative outperformance to the standard DNS model, whereas bold values indicate significant outperformance at a 95% significance level according to the Diebold-Mariano test.

Table 9: Root mean squared prediction errors, 2003/10-2010/5

Model / Maturity	3	6	12	24	36	60	84	120	240	360
Panel A: 1-month horizon										
DNS	38.5	27.4	24.6	29.4	30.5	28.2	27.9	26.1	29.5	28.6
DNS-L	**0.98**	**0.98**	**0.96**	0.97	0.99	1.02	1.02	1.00	0.99	**0.90**
DNS-V	0.99	0.99	1.06	1.33	1.63	2.01	2.04	2.09	1.29	1.71
DNS-M(ES)	**0.96**	1.00	1.02	1.01	1.00	0.98	**0.98**	**0.98**	1.04	1.00
DNS-M(PCA)	0.97	0.97	1.04	1.02	1.01	1.01	1.02	1.04	1.09	1.07
DNS-M(PLS)	0.94	0.99	1.07	1.02	1.01	1.04	1.06	1.09	1.11	1.14
DNS-LVM(ES)	**0.90**	**0.86**	**0.88**	0.95	1.04	1.18	1.22	1.22	0.95	1.09
DNS-LVM(PCA)	0.96	0.93	1.06	1.14	1.31	1.56	1.62	1.69	1.07	1.45
DNS-LVM(PLS)	0.94	0.94	1.04	1.09	1.25	1.47	1.53	1.57	1.08	1.29
Panel B: 3-month horizon										
DNS	69.1	61.6	59.1	61.7	60.9	52.7	49.2	43.6	36.7	41.6
DNS-L	0.99	0.98	**0.96**	**0.95**	**0.95**	0.96	0.97	0.97	1.01	0.97
DNS-V	0.98	**0.96**	**0.93**	1.01	1.12	1.32	1.38	1.46	1.25	1.38
DNS-M(ES)	0.96	1.00	1.00	1.00	0.98	0.97	0.95	0.95	1.03	0.97
DNS-M(PCA)	0.97	0.96	0.98	1.00	1.01	1.00	1.01	1.01	1.07	1.03
DNS-M(PLS)	0.91	0.93	0.95	0.96	0.97	1.00	1.04	1.07	1.14	1.14
DNS-LVM(ES)	0.85	0.84	0.86	0.92	0.96	1.04	1.07	1.08	1.08	1.06
DNS-LVM(PCA)	0.94	0.91	0.92	0.96	1.03	1.16	1.23	1.29	1.12	1.23
DNS-LVM(PLS)	0.91	0.89	0.88	0.92	0.98	1.08	1.15	1.18	1.09	1.11
Panel C: 6-month horizon										
DNS	110.5	103.8	98.5	94.7	89.2	76.0	68.2	59.7	44.0	52.7
DNS-L	0.99	0.99	0.97	**0.96**	**0.95**	**0.94**	0.95	0.97	1.05	1.02
DNS-V	0.98	0.97	**0.94**	**0.95**	1.01	1.12	1.18	1.25	1.24	1.27
DNS-M(ES)	0.97	1.00	1.00	0.99	0.97	0.94	0.92	**0.91**	0.98	**0.91**
DNS-M(PCA)	0.93	0.94	0.96	0.98	0.99	1.00	1.00	1.00	1.01	0.99
DNS-M(PLS)	0.90	0.89	0.90	0.92	0.95	1.02	1.09	1.12	1.24	1.20
DNS-LVM(ES)	0.85	0.86	0.87	0.91	0.93	0.98	1.02	1.03	1.15	1.07
DNS-LVM(PCA)	0.93	0.92	0.91	0.92	0.95	1.04	1.11	1.16	1.17	1.19
DNS-LVM(PLS)	**0.87**	**0.87**	**0.85**	**0.85**	0.87	0.94	1.01	1.04	1.12	1.07
Panel D: 12-month horizon										
DNS	190.5	180.8	168.3	152.1	137.1	110.5	95.3	82.5	58.8	71.9
DNS-L	1.00	1.00	0.99	0.97	0.95	0.91	0.90	0.90	0.95	0.97
DNS-V	0.99	0.98	**0.96**	**0.95**	**0.96**	1.02	1.06	1.10	1.15	1.13
DNS-M(ES)	0.98	0.99	0.99	0.98	0.96	0.92	0.88	0.85	0.85	0.81
DNS-M(PCA)	0.93	0.93	0.94	0.96	0.97	0.98	0.99	0.99	1.00	1.00
DNS-M(PLS)	**0.92**	**0.91**	**0.91**	**0.93**	**0.95**	1.02	1.09	1.14	1.29	1.21
DNS-LVM(ES)	0.94	0.93	0.91	0.89	0.86	0.83	0.83	0.82	0.94	0.87
DNS-LVM(PCA)	1.06	1.03	0.99	0.96	0.96	0.99	1.04	1.08	1.14	1.13
DNS-LVM(PLS)	0.89	0.88	0.84	0.81	0.81	0.83	0.88	0.92	1.06	1.00

NOTE: This table reports the forecast performance of all models described in section 2, for the period 2003/10 to 2010/5. For further details see table 8.

Table 10: Root mean squared prediction errors, 2010/6-2017/1

Model / Maturity	3	6	12	24	36	60	84	120	240	360
Panel A: 1-month horizon										
DNS	17.1	10.7	10.1	15.3	20.5	25.8	25.7	24.5	26.7	27.8
DNS-L	**0.79**	**0.80**	**0.69**	**0.75**	**0.86**	**0.87**	**0.89**	**0.94**	0.97	**0.90**
DNS-V	**0.87**	1.17	1.54	2.19	2.58	2.78	2.93	2.94	2.56	2.23
DNS-M(ES)	**0.95**	**0.91**	**0.84**	**0.88**	**0.89**	**0.93**	0.95	0.96	**0.96**	1.02
DNS-M(PCA)	1.11	1.18	1.15	1.08	1.05	1.02	1.02	1.01	1.01	1.00
DNS-M(PLS)	1.20	1.34	1.29	1.15	1.09	1.03	1.01	1.00	0.99	1.01
DNS-LVM(ES)	**0.60**	0.84	**0.75**	**0.73**	**0.80**	**0.82**	**0.88**	0.93	**0.88**	**0.85**
DNS-LVM(PCA)	**0.63**	**0.64**	**0.65**	**0.77**	**0.91**	0.98	1.04	1.09	1.10	1.06
DNS-LVM(PLS)	**0.63**	**0.71**	**0.74**	0.89	1.03	1.08	1.12	1.15	1.14	1.07
Panel B: 3-month horizon										
DNS	31.6	26.5	26.4	33.1	42.6	50.8	51.0	50.5	52.5	50.0
DNS-L	**0.77**	**0.75**	**0.72**	**0.74**	**0.79**	**0.84**	**0.87**	**0.89**	**0.93**	**0.93**
DNS-V	**0.66**	**0.74**	0.98	1.35	1.51	1.64	1.68	1.64	1.50	1.45
DNS-M(ES)	**0.92**	**0.89**	**0.82**	**0.86**	**0.90**	0.93	0.94	0.94	0.94	0.98
DNS-M(PCA)	1.13	1.14	1.12	1.06	1.03	1.01	1.00	1.00	1.00	0.99
DNS-M(PLS)	1.32	1.38	1.36	1.21	1.11	1.04	1.02	1.00	0.97	0.98
DNS-LVM(ES)	**0.45**	**0.51**	**0.47**	**0.55**	**0.64**	**0.75**	0.83	0.88	**0.85**	**0.87**
DNS-LVM(PCA)	**0.50**	**0.46**	**0.46**	**0.60**	**0.70**	**0.81**	0.88	0.91	0.92	0.94
DNS-LVM(PLS)	**0.48**	**0.51**	**0.60**	**0.78**	**0.87**	0.94	0.99	1.00	0.98	1.00
Panel C: 6-month horizon										
DNS	50.4	46.6	48.3	57.2	68.7	77.8	77.5	76.7	77.8	71.9
DNS-L	**0.78**	**0.76**	**0.74**	**0.73**	**0.76**	**0.80**	**0.82**	**0.85**	**0.89**	**0.90**
DNS-V	**0.59**	**0.65**	**0.80**	1.05	1.18	1.28	1.32	1.28	1.21	1.20
DNS-M(ES)	**0.89**	**0.87**	**0.84**	0.87	0.90	0.92	0.93	0.93	0.92	0.96
DNS-M(PCA)	1.09	1.09	1.07	1.03	1.01	1.00	0.99	0.99	0.99	0.99
DNS-M(PLS)	1.35	1.36	1.31	1.18	1.10	1.04	1.01	0.99	**0.96**	**0.96**
DNS-LVM(ES)	**0.41**	**0.43**	**0.36**	**0.40**	**0.52**	**0.69**	0.78	0.84	**0.82**	**0.84**
DNS-LVM(PCA)	**0.42**	**0.40**	**0.38**	**0.49**	**0.59**	**0.71**	**0.78**	**0.81**	**0.84**	**0.87**
DNS-LVM(PLS)	**0.42**	**0.47**	**0.55**	**0.70**	**0.78**	**0.87**	0.91	0.92	0.90	0.94
Panel D: 12-month horizon										
DNS	79.2	78.1	83.7	96.5	109.0	116.5	114.0	111.0	109.6	100.3
DNS-L	**0.83**	**0.81**	**0.78**	**0.76**	**0.76**	**0.77**	**0.79**	**0.81**	**0.87**	**0.88**
DNS-V	**0.59**	**0.64**	**0.75**	**0.92**	1.01	1.09	1.11	1.10	1.07	1.08
DNS-M(ES)	**0.88**	**0.88**	0.87	0.90	0.92	0.93	0.93	0.91	0.89	0.91
DNS-M(PCA)	1.04	1.03	1.02	1.00	0.99	0.99	0.99	0.99	0.99	0.99
DNS-M(PLS)	1.31	1.29	1.21	1.12	1.07	1.03	1.01	0.99	0.97	0.97
DNS-LVM(ES)	**0.39**	**0.37**	**0.30**	**0.25**	**0.34**	**0.51**	0.61	0.69	**0.71**	**0.72**
DNS-LVM(PCA)	**0.38**	**0.35**	**0.30**	**0.35**	**0.44**	**0.57**	**0.64**	**0.69**	**0.77**	0.81
DNS-LVM(PLS)	**0.41**	**0.47**	**0.54**	**0.66**	**0.73**	**0.83**	**0.87**	**0.87**	**0.86**	0.90

NOTE: This table reports the forecast performance of all models described in section 2, for the period 2010/6 to 2017/1. for further details see table 8.

6 Conclusion

In this study we investigate the in- and out-of-sample performance of several extensions in the DNS model proposed by Diebold et al. (2006). We consider extensions that allow for time variation in the loading parameter, volatility and unconditional mean. Furthermore, we combine these extensions into more sophisticated models. The purpose of these extensions is twofold. First, time variation in the parameters makes the model more flexible such that it is better able to reproduce yield curve dynamics. Second, with the extension of a time-varying unconditional mean we include macroeconomic factors to exploit macroeconomic information. We use a dataset of the United Stated Department of Treasury containing yields with a maturity up to 30 years, which is longer than considered in most studies. Moreover, the dataset spans the time period from October 1993 trough January 2017 and thereby contains a recent period that has not been used in many studies yet.

All extensions lead to significant improvement in in-sample fit, according to the likelihood ratio statistics. However, gains are more pronounced for the extensions with time variation in the loading parameter and volatility than in the unconditional mean. Combining extensions into one model results in even more significant improvement in the model fit.

With comparing the predictive performance of the extended models with the standard DNS model over two periods, we find that the performance of extensions depends heavily on the the period and forecast horizon. In general, extended models perform better for longer horizons. Models without macroeconomic information perform better compared to the DNS model in a calm period such as 2010/6-2017/1 than in a volatile period such as 2003/6-2017/1, whereas models with macroeconomic information perform better in a volatile period. This may be because usually more economically important events occur in times of high volatility and therefore macroeconomic variables possibly contain more useful information. The extension of a time-varying loading parameter and the extension of a time-varying unconditional mean without macroeconomic information improve on the DNS model for almost all maturities at all horizons for both calm and volatile periods, whereas the model with time-varying volatility improves only for the short maturities. Extensions with macroeconomic information are only beneficial for maturities and periods with high volatility. For these cases, the model with factors constructed using PLS provides better forecasts than the model with factors constructed using PCA. Combining extensions can lead to substantial improvement in out-of-sample performance compared the models with individual extensions. In particular when extensions that perform well in a particular period are combined, this leads to substantial improvement compared to all single-extension models in this period.

Finally, from comparing in- and out-of-sample performance for each model we can conclude that strong improvements in in-sample fit does not always lead to strong improvements in out-of-sample performance and that relatively small in-sample improvement does not always imply small improvement in out-of-sample performance. This contradictory performance is especially visible for the DNS-V and DNS-M models.

7 Further Research

The models and estimation methods we discussed during this study leave great space for further research and improvement. Below, we briefly discuss several directions for further research. First, the errors in the measurement equation as well as the state equation are assumed to be Gaussian distributed in this study. However, this assumption may be inappropriate and a distribution with heavier tails may be better suitable. Hence, it would be interesting to consider a model with heavier tails, e.g. use a t-distribution or a mixture of normal distributions for the error terms. Second, in an attempt to make the models more flexible one can introduce a fourth latent factor as in the Svensson (1995) model. It would be interesting to see how the earlier described extensions would perform in the presence of an additional latent factor. Third, in this study we introduced the rather simple GARCH specification for the volatility component in the DNS-V model. One may opt for a more advanced specification such as T-GARCH(X) or introducing varying volatility in different manners to the DNS model. Fourth, with the introduction of the extensions the number of parameters in the models has increased. Because of the higher number of parameters, the parameters contain a higher level of uncertainty. It may be therefore useful to consider Bayesian estimation methods that take parameter uncertainty into account. This is for example done by De Pooter et al. (2010) and they obtain encouraging results. Finally, we used the extended Kalman filter for estimating the latent factors in non-linear models, following Koopman et al. (2010). However, several other options are possible to estimate the latent factors in the non-linear model. One option would be to use the more recently developed unscented Kalman filter (Julier & Uhlmann, 2004). Besides this, other options can be found in simulation based filtering algorithms.

References

Bianchi, F., Mumtaz, H., & Surico, P. (2006). *The uk great stability: a view from the term structure of interest rates* (Tech. Rep.). Working Paper.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542–547.

Christensen, B. J., van der Wel, M., et al. (2010). *An asset pricing approach to testing general term structure models including heath-jarrow-morton specifications and affine subclasses* (Tech. Rep.). Department of Economics and Business Economics, Aarhus University.

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, *164*(1), 4–20.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 385–407.

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*.

De Pooter, M., Ravazzolo, F., & Van Dijk, D. J. (2010). Term structure forecasting using macro factors and forecast combination.

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

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 253–263.

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

Dijk, D., Koopman, S. J., Wel, M., & Wright, J. H. (2014). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*, *29*(5), 693–712.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (No. 38). Oxford University Press.

Exterkate, P., Dijk, D. V., Heij, C., & 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.

Garthwaite, P. H. (1994). An interpretation of partial least squares. *Journal of the American Statistical Association*, *89*(425), 122–127.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

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

Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, *92*(3), 401–422.

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

McCracken, M. W., & Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, *146*(1), 26–43.

Mönch, E. (2012). Term structure surprises: the predictive content of curvature, level, and slope. *Journal of Applied Econometrics*, *27*(4), 574–602.

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

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Svensson, L. E. (1995). Estimating forward interest rates with the extended nelson & siegel method. *Sveriges Riksbank Quarterly Review*, *3*(1), 13–26.

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

Yu, W.-C., & Zivot, E. (2011). Forecasting the term structures of treasury and corporate yields using dynamic nelson-siegel models. *International Journal of Forecasting*, *27*(2), 579–591.

Appendix

A Macroeconomic Dataset

The macroeconomic dataset is composed by McCracken and Ng (2016) and is designed for empirical analysis of big data. This appendix is a copy of the appendix in their paper.

The column TCODE denotes the following transformation for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$; (7) $\Delta(x_t/x_{t-1} - 1.0)$. The 'fred' column gives mnemonics in the Federal Reserve Database followed by a short description. The comparable series in the Global Insight Database is given in the column GSI. All ISM series are excluded from the dataset, since they are no longer updated after June 2016.

Some series require adjustments to the raw data available in FRED. These variables are tagged with an asterisk to indicate that they have been adjusted and thus differ from the series from the source. A summary of the adjustments is detailed in the paper https://research.stlouisfed.org/wp/2015/2015-012.pdf.

Group 1: Output and income

	id	tcode	fred	description	gsi	gsi:description
1	1	5	RPI	Real Personal Income	M_14386177	PI
2	2	5	W875RX1	Real personal income ex transfer receipts	M_145256755	PI less transfers
3	6	5	INDPRO	IP Index	M_116460980	IP: total
4	7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	M_116460981	IP: products
5	8	5	IPFINAL	IP: Final Products (Market Group)	M_116461268	IP: final prod
6	9	5	IPCONGD	IP: Consumer Goods	M_116460982	IP: cons gds
7	10	5	IPDCONGD	IP: Durable Consumer Goods	M_116460983	IP: cons dble
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods	M_116460988	IP: cons nondble
9	12	5	IPBUSEQ	IP: Business Equipment	M_116460995	IP: bus eqpt
10	13	5	IPMAT	IP: Materials	M_116461002	IP: matls
11	14	5	IPDMAT	IP: Durable Materials	M_116461004	IP: dble matls
12	15	5	IPNMAT	IP: Nondurable Materials	M_116461008	IP: nondble matls
13	16	5	IPMANSICS	IP: Manufacturing (SIC)	M_116461013	IP: mfg
14	17	5	IPB51222s	IP: Residential Utilities	M_116461276	IP: res util
15	18	5	IPFUELS	IP: Fuels	M_116461275	IP: fuels
16	19	1	NAPMPI	ISM Manufacturing: Production Index	M_110157212	NAPM prodn
17	20	2	CUMFNS	Capacity Utilization: Manufacturing	M_116461602	Cap util

Group 2: Labor market

	id	tcode	fred	description	gsi	gsi:description
1	21*	2	HWI	Help-Wanted Index for United States		Help wanted indx
2	22*	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	M_110156531	Help wanted/unemp
3	23	5	CLF16OV	Civilian Labor Force	M_110156467	Emp CPS total
4	24	5	CE16OV	Civilian Employment	M_110156498	Emp CPS nonag
5	25	2	UNRATE	Civilian Unemployment Rate	M_110156541	U: all
6	26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)	M_110156528	U: mean duration
7	27	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	M_110156527	U < 5 wks
8	28	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	M_110156523	U 5-14 wks
9	29	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	M_110156524	U 15+ wks
10	30	5	UEMP15T26	Civilians Unemployed for 15-26 Weeks	M_110156525	U 15-26 wks
11	31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	M_110156526	U 27+ wks
12	32*	5	CLAIMSx	Initial Claims	M_15186204	UI claims
13	33	5	PAYEMS	All Employees: Total nonfarm	M_123109146	Emp: total
14	34	5	USGOOD	All Employees: Goods-Producing Industries	M_123109172	Emp: gds prod
15	35	5	CES1021000001	All Employees: Mining and Logging: Mining	M_123109244	Emp: mining
16	36	5	USCONS	All Employees: Construction	M_123109331	Emp: const
17	37	5	MANEMP	All Employees: Manufacturing	M_123109542	Emp: mfg
18	38	5	DMANEMP	All Employees: Durable goods	M_123109573	Emp: dble gds
19	39	5	NDMANEMP	All Employees: Nondurable goods	M_123110741	Emp: nondbles
20	40	5	SRVPRD	All Employees: Service-Providing Industries	M_123109193	Emp: services
21	41	5	USTPU	All Employees: Trade, Transportation & Utilities	M_123111543	Emp: TTU
22	42	5	USWTRADE	All Employees: Wholesale Trade	M_123111563	Emp: wholesale
23	43	5	USTRADE	All Employees: Retail Trade	M_123111867	Emp: retail
24	44	5	USFIRE	All Employees: Financial Activities	M_123112777	Emp: FIRE
25	45	5	USGOVT	All Employees: Government	M_123114411	Emp: Govt
26	46	1	CES0600000007	Avg Weekly Hours : Goods-Producing	M_140687274	Avg hrs
27	47	2	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	M_123109554	Overtime: mfg
28	48	1	AWHMAN	Avg Weekly Hours : Manufacturing	M_14386098	Avg hrs: mfg
29	49	1	NAPMEI	ISM Manufacturing: Employment Index	M_110157206	NAPM empl
30	127	6	CES0600000008	Avg Hourly Earnings : Goods-Producing	M_123109182	AHE: goods
31	128	6	CES2000000008	Avg Hourly Earnings : Construction	M_123109341	AHE: const
32	129	6	CES3000000008	Avg Hourly Earnings : Manufacturing	M_123109552	AHE: mfg

Group 3: Housing

	id	tcode	fred	description	gsi	gsi:description
1	50	4	HOUST	Housing Starts: Total New Privately Owned	M_110155536	Starts: nonfarm
2	51	4	HOUSTNE	Housing Starts, Northeast	M_110155538	Starts: NE
3	52	4	HOUSTMW	Housing Starts, Midwest	M_110155537	Starts: MW
4	53	4	HOUSTS	Housing Starts, South	M_110155543	Starts: South
5	54	4	HOUSTW	Housing Starts, West	M_110155544	Starts: West
6	55	4	PERMIT	New Private Housing Permits (SAAR)	M_110155532	BP: total
7	56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)	M_110155531	BP: NE
8	57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)	M_110155530	BP: MW
9	58	4	PERMITS	New Private Housing Permits, South (SAAR)	M_110155533	BP: South
10	59	4	PERMITW	New Private Housing Permits, West (SAAR)	M_110155534	BP: West

Group 4: Consumption, orders, and inventories

	id	tcode	fred	description	gsi	gsi:description
1	3	5	DPCERA3M086SBEA	Real personal consumption expenditures	M_123008274	Real Consumption
2	4*	5	CMRMTSPLx	Real Manu. and Trade Industries Sales	M_110156998	M&T sales
3	5*	5	RETAILx	Retail and Food Services Sales	M_130439509	Retail sales
4	60	1	NAPM	ISM : PMI Composite Index	M_110157208	PMI
5	61	1	NAPMNOI	ISM : New Orders Index	M_110157210	NAPM new ordrs
6	62	1	NAPMSDI	ISM : Supplier Deliveries Index	M_110157205	NAPM vendor del
7	63	1	NAPMII	ISM : Inventories Index	M_110157211	NAPM Invent
8	64	5	ACOGNO	New Orders for Consumer Goods	M_14385863	Orders: cons gds
9	65*	5	AMDMNOx	New Orders for Durable Goods	M_14386110	Orders: dble gds
10	66*	5	ANDENOx	New Orders for Nondefense Capital Goods	M_178554409	Orders: cap gds
11	67*	5	AMDMUOx	Unfilled Orders for Durable Goods	M_14385946	Unf orders: dble
12	68*	5	BUSINVx	Total Business Inventories	M_15192014	M&T invent
13	69*	2	ISRATIOx	Total Business: Inventories to Sales Ratio	M_15191529	M&T invent/sales
14	130*	2	UMCSENTx	Consumer Sentiment Index	hhsntn	Consumer expect

Group 5: Money and credit

	id	tcode	fred	description	gsi	gsi:description
1	70	6	M1SL	M1 Money Stock	M_110154984	M1
2	71	6	M2SL	M2 Money Stock	M_110154985	M2
3	72	5	M2REAL	Real M2 Money Stock	M_110154985	M2 (real)
4	73	6	AMBSL	St. Louis Adjusted Monetary Base	M_110154995	MB
5	74	6	TOTRESNS	Total Reserves of Depository Institutions	M_110155011	Reserves tot
6	75	7	NONBORRES	Reserves Of Depository Institutions	M_110155009	Reserves nonbor
7	76	6	BUSLOANS	Commercial and Industrial Loans	BUSLOANS	C&I loan plus
8	77	6	REALLN	Real Estate Loans at All Commercial Banks	BUSLOANS	DC&I loans
9	78	6	NONREVSL	Total Nonrevolving Credit	M_110154564	Cons credit
10	79*	2	CONSPI	Nonrevolving consumer credit to Personal Income	M_110154569	Inst cred/PI
11	131	6	MZMSL	MZM Money Stock	N.A.	N.A.
12	132	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	N.A.	N.A.
13	133	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding	N.A.	N.A.
14	134	6	INVEST	Securities in Bank Credit at All Commercial Banks	N.A.	N.A.

Group 6: Interest and exchange rates

	id	tcode	fred	description	gsi	gsi:description
1	84	2	FEDFUNDS	Effective Federal Funds Rate	M_110155157	Fed Funds
2	85*	2	CP3Mx	3-Month AA Financial Commercial Paper Rate	CPF3M	Comm paper
3	86	2	TB3MS	3-Month Treasury Bill:	M_110155165	3 mo T-bill
4	87	2	TB6MS	6-Month Treasury Bill:	M_110155166	6 mo T-bill
5	88	2	GS1	1-Year Treasury Rate	M_110155168	1 yr T-bond
6	89	2	GS5	5-Year Treasury Rate	M_110155174	5 yr T-bond
7	90	2	GS10	10-Year Treasury Rate	M_110155169	10 yr T-bond
8	91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield		Aaa bond
9	92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield		Baa bond
10	93*	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS		CP-FF spread
11	94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS		3 mo-FF spread
12	95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS		6 mo-FF spread
13	96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS		1 yr-FF spread
14	97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS		5 yr-FF spread
15	98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS		10 yr-FF spread
16	99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS		Aaa-FF spread
17	100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS		Baa-FF spread
18	101	5	TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies		Ex rate: avg
19	102*	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	M_110154768	Ex rate: Switz
20	103*	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	M_110154755	Ex rate: Japan
21	104*	5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	M_110154772	Ex rate: UK
22	105*	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	M_110154744	EX rate: Canada

Group 7: Prices

	id	tcode	fred	description	gsi	gsi:description
1	106	6	WPSFD49207	PPI: Finished Goods	M110157517	PPI: fin gds
2	107	6	WPSFD49502	PPI: Finished Consumer Goods	M110157508	PPI: cons gds
3	108	6	WPSID61	PPI: Intermediate Materials	M_110157527	PPI: int matls
4	109	6	WPSID62	PPI: Crude Materials	M_110157500	PPI: crude matls
5	110*	6	OILPRICEx	Crude Oil, spliced WTI and Cushing	M_110157273	Spot market price
6	111	6	PPICMM	PPI: Metals and metal products:	M_110157335	PPI: nonferrous
7	112	1	NAPMPRI	ISM Manufacturing: Prices Index	M_110157204	NAPM com price
8	113	6	CPIAUCSL	CPI : All Items	M_110157323	CPI-U: all
9	114	6	CPIAPPSL	CPI : Apparel	M_110157299	CPI-U: apparel
10	115	6	CPITRNSL	CPI : Transportation	M_110157302	CPI-U: transp
11	116	6	CPIMEDSL	CPI : Medical Care	M_110157304	CPI-U: medical
12	117	6	CUSR0000SAC	CPI : Commodities	M_110157314	CPI-U: comm.
13	118	6	CUUR0000SAD	CPI : Durables	M_110157315	CPI-U: dbles
14	119	6	CUSR0000SAS	CPI : Services	M_110157325	CPI-U: services
15	120	6	CPIULFSL	CPI : All Items Less Food	M_110157328	CPI-U: ex food
16	121	6	CUUR0000SA0L2	CPI : All items less shelter	M_110157329	CPI-U: ex shelter
17	122	6	CUSR0000SA0L5	CPI : All items less medical care	M_110157330	CPI-U: ex med
18	123	6	PCEPI	Personal Cons. Expend.: Chain Index	gmdc	PCE defl
19	124	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	gmdcd	PCE defl: dlbes
20	125	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	gmdcn	PCE defl: nondble
21	126	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	gmdcs	PCE defl: service

Group 8: Stock market

	id	tcode	fred	description	gsi	gsi:description
1	80*	5	S&P 500	S&P's Common Stock Price Index: Composite	M_110155044	S&P 500
2	81*	5	S&P: indust	S&P's Common Stock Price Index: Industrials	M_110155047	S&P: indust
3	82*	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield		S&P div yield
4	83*	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio		S&P PE ratio
5	135*	1	VXOCLSx	VXO		

B Unconditional Covariance Matrix of the State Vector

In order to solve $\Sigma_x - A\Sigma_x A' = \Sigma_\nu$ we follow the procedure of B. J. Christensen, van der Wel, et al. (2010). First, we vectorize both sides of the equation such that we obtain

$$\mathrm{vec}(\Sigma_x) - \mathrm{vec}(A\Sigma_x A) = \mathrm{vec}(\Sigma_\nu).$$

Next, we rewrite this equation as

$$I_{d^2} \cdot \mathrm{vec}(\Sigma_x) - (A \otimes A) \cdot \mathrm{vec}(\Sigma_x) = (I_{d^2} - (A \otimes A)) \cdot \mathrm{vec}(\Sigma_x) = \mathrm{vec}(\Sigma_\nu),$$

where d is the dimension of x_t. Now, the above can easily be solved. This results in the following solution

$$\mathrm{vec}(\Sigma_x) = \left(I_{d^2} - A \otimes A'\right)^{-1} \cdot \mathrm{vec}(\Sigma_\nu)$$

C Kalman Filter: Derivation

Consider the general state space representation in (4) and (5). We assume that the parameters in these specifications are known. To explain the Kalman filter, we first consider the forecast of x_t conditional on the information known at time $t-1$, i.e.

$$\hat{x}_{t|t-1} = \mathrm{E}\left[x_t | \mathcal{I}_t\right], \tag{45}$$

where $\mathcal{I}_{t-1}$ is the information set up to time t. This conditional expectation of x_{t+1} is the optimal forecast given the information up to time t. Moreover, we are interested in the conditional variance of x_t. This is given by

$$P_{t+1|t} = \mathrm{E}\left[\left(x_{t+1} - \hat{x}_{t+1|t}\right)\left(x_{t+1} - \hat{x}_{t+1|t}\right)' \Big| \mathcal{I}_{t-1}\right]. \tag{46}$$

Now, suppose that we have obtained the optimal forecasts $\hat{x}_{t|t-1}$ and the corresponding variance $P_{t|t-1}$. Next, we are interested in constructing the optimal forecast $\hat{x}_{t+1|t}$ and the corresponding variance $P_{t+1|t}$ for the next observation. If the state variable x_t was observable this would, with use of the state equation (5), result in

$$\hat{x}_{t+1|t} = C + Ax_t \quad \text{and} \quad P_{t+1|t} = Q. \tag{47}$$

However, x_t is not observed and we end up with

$$\begin{aligned} \hat{x}_{t+1|t} &= \mathrm{E}\left[x_{t+1} | \mathcal{I}_t\right] \\ &= \mathrm{E}\left[C + Ax_t + \nu_t | \mathcal{I}_t\right] \\ &= C + A \cdot \mathrm{E}\left[x_t | \mathcal{I}_t\right]. \end{aligned}$$

Therefore, we first need to obtain $\hat{x}_{t|t} \equiv \mathrm{E}[x_t|\mathcal{I}_t]$. Now, this is the point where the Kalman filter comes in. The Kalman filter is based on the property of two variables that have a joint normal distribution

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}\right). \tag{48}$$

Then, the distribution of β_t conditional on y_t is $\mathcal{N}(m, \Sigma)$, with

$$m = \mu_2 + \Omega_{21}\Omega_{11}^{-1}(z_1 - \mu_1) \tag{49}$$

$$\Sigma = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}. \tag{50}$$

The Kalman filter applies this results to each observation $t = 1, \dots, T$ sequentially. In our case we are looking for the joint distribution of y_t and x_t conditional on $\mathcal{I}_{t-1}$. However, in order to use this result, the conditional means of y_t and β_t and their conditional variances and covariances are required. The conditional means follow directly from (4) and (5). For the conditional variances and covariances consider the forecast errors

$$\begin{aligned} y_t - \mathrm{E}[y_t|\mathcal{I}_{t-1}] &= (Hx_t + \omega_t) - \left(H\hat{x}_{t|t-1}\right) \\ &= H\left(\beta_t - b_{t|t-1}\right) + \omega_t. \end{aligned} \tag{51}$$

whose variance is the same as the variance of y_t given x_t and $\mathcal{I}_{t-1}$ and equals

$$\mathrm{V}[y_t|\mathcal{I}_{t-1}] = \mathrm{V}[y_t - \mathrm{E}[y_t|\mathcal{I}_{t-1}]] = HP_{t|t-1}H' + R. \tag{52}$$

Next, the covariance equals

$$\begin{aligned} \mathrm{Cov}[y_t, x_t|\mathcal{I}_{t-1}] &= \mathrm{E}\left[\left(H'\left(x_t - \hat{x}_{t|t-1}\right) + \omega_t\right)\left(x_t - \hat{x}_{t|t-1}\right)\middle|\,\mathcal{I}_{t-1}\right] \\ &= HP_{t|t-1} \end{aligned} \tag{53}$$

Now, all required components are available and the results from (45), (52) and (53) can be used to specify the joint distribution of y_t and x_t conditional on x_t and $\mathcal{I}_{t-1}$:

$$\left.\begin{pmatrix} y_t \\ x_t \end{pmatrix}\right|\mathcal{I}_{t-1} \sim \mathcal{N}\left[\begin{pmatrix} H'\hat{x}_{t|t-1} \\ \hat{x}_{t|t-1} \end{pmatrix}, \begin{pmatrix} HP_{t|t-1}H' + R & HP_{t|t-1} \\ P_{t|t-1}H' & P_{t|t-1} \end{pmatrix}\right] \tag{54}$$

By using the result in (49) and (50) it follows that

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t v_t, \tag{55}$$

$$P_{t|t} = P_{t|t-1} - K_t H' P_{t|t-1}, \tag{56}$$

where $v_t = y_t - H'\hat{x}_{t|t-1}$ is the prediction error, $K_t = P_{t|t-1}H'F_t^{-1}$ is called the Kalman gain and $F_t = HP_{t|t-1}H' + R$ is the measurement prediction variance. This step is often referred to as the filtering step. When substituting the estimates from (55) in the state equation in (5), it follows that

$$\hat{x}_{t+1|t} = C + A\hat{x}_{t|t} \tag{57}$$

$$P_{t+1|t} = AP_{t|t}A' + Q. \tag{58}$$

Combining these expressions with (55) and (56), we obtain the desired results and the recursion is complete. This is called the prediction step. Finally, we need starting values $\hat{x}_{1|0}$ and $P_{1|0}$ in order to start the recursion. A common choice for this purpose is to use the unconditional mean and variance of x_t.

D Kalman Filter: Univariate Treatment of Multivariate Series

In section 2.7 we have treated the filtering of the multivariate time series in the traditional way by taking the entire observational vectors y_t as items for analysis. In this appendix we describe the alternative approach, explained by Durbin and Koopman (2012), where the elements of y_t are brought into the filtering process one at a time. The main advantage of this approach is that it offers computational gains and resolves computational problems.

The analysis will be based on the general model described by (4) and (5) in which we assume that Q and R are diagonal. To begin with, we introduce some new notation. Write the observation and disturbance vectors as

$$H = \begin{bmatrix} H_1 \\ \vdots \\ H_m \end{bmatrix}, \qquad R = \begin{bmatrix} \sigma^2_{t,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{t,p} \end{bmatrix}$$

where $y_{t,i}, \varepsilon_{t,i}$ and $\sigma^2_{t,i}$ are scalars and H_i is a $1 \times p$ row vector, for $i = 1, \cdots, p$. Now, consider the observation equation for the univariate representation of the model, this is given by

$$y_{t,i} = H_i x_{t,i} + \varepsilon_{t,i}, \tag{59}$$

where $x_{t,i} = x_t$ and $x_{t,i}$ is not the i-th element of x_t but the predicted state vector after taking observation i into account. The corresponding state equation is then

$$\begin{aligned} x_{t,i+1} &= x_{t,i}, & i &= 1, \cdots, p-1 \\ x_{t+1,1} &= Ax_{t,i} + \nu_{t+1}, & t &= 1, \cdots, T \end{aligned} \tag{60}$$

where the initial state vector $x_{1,1} = x_1 \sim \mathcal{N}(x_1, P_1)$. Next, define

$$x_{t,i} = \mathrm{E}[x_{t,1}|Y_{t-1}], \qquad P_{t,1} = \mathrm{Var}[x_{t,1}|Y_{t-1}]$$

and

$$x_{t,i} = \mathrm{E}[x_{t,i}|Y_{t-1}, y_{t,1}, \cdots, y_{t,i-1}], \qquad P_{t,i} = \mathrm{Var}[x_{t,i}|Y_{t-1}, y_{t,1}, \cdots, y_{t,i-1}],$$

for $i = 2, \cdots, p$. The key aspect in this approach is to treat the vector series $y_1, \cdots, y_T$ as the scalar series

$$y_{1,1}, \cdots, y_{1,p}, y_{2,1}, \cdots, y_{T,p}.$$

This makes that the filtering equations can be written as

$$\begin{aligned} \hat{x}_{t|t,i+1} &= \hat{x}_{t|t-1,i} + K_i v_{t,i}, \\ P_{t|t,i+1} &= P_{t|t-1,i} - K_i H_i P_{t|t-1,i}, \end{aligned} \tag{61}$$

where $v_{t,i} = y_{t,i} - H_i x_{t,i}$ and $K_{t,i} = P_{t|t-1,i} H_i' (H_i P_{t|t-1,i} H_i' + \sigma^2_{t,i})$, for $i = 1, \cdots, p$ and $t = 1, \cdots, T$. In this formulation $v_{t,i}$ is a scalar and $K_{t,i}$ is a column vector of size $p \times 1$. The update of the state vector and its covariance matrix is based on the filtered estimation at time t, where the information of all p observations is taken into account. Hence, the prediction step is given by the following equations

$$\begin{aligned} x_{t+1|t,i} &= A x_{t|t,p} \\ P_{t+1|t,i} &= A P_{t,p} A' + Q \end{aligned} \tag{62}$$

for $i = 1, \cdots, p$. The values of $x_{t+1|t}$ and $P_{t+1|t}$ are the same as in the standard Kalman filter. However, it is important to note that this is not the case for the elements of the innovation vector v_t and $v_{t,i}$, only the first element of v_t is equal $v_{t,1}$. The same holds for the variance matrix of y_t, where only the first diagonal element is equal. For further information about the univariate treatment of a multivariate time series using the Kalman filter we refer to Durbin and Koopman (2012).

E Additional Tables

Table 11: Cross-correlations of yields

	3	6	12	24	36	60	84	120	240	360
3	1.000	0.998	0.993	0.978	0.964	0.935	0.910	0.880	0.829	0.806
6		1.000	0.997	0.985	0.973	0.944	0.920	0.890	0.838	0.814
12			1.000	0.994	0.985	0.961	0.939	0.913	0.862	0.841
24				1.000	0.997	0.982	0.966	0.943	0.897	0.878
36					1.000	0.993	0.981	0.961	0.921	0.903
60						1.000	0.997	0.986	0.956	0.942
84							1.000	0.995	0.975	0.963
120								1.000	0.990	0.983
240									1.000	0.995
360										1.000

NOTE: This table reports the cross-correlations of the yields over the period October 1994 trough January 2017. Maturities are shown in months.

Table 12: AR(1) estimates of the DNS-V model.

(a) DNS-V coefficient matrix and constant

	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$	μ
$\beta_{1,t}$ (level)	0.986^{***} (0.0090)	0	0	5.114^{***} (0.9235)
$\beta_{2,t}$ (slope)	0	0.983^{***} (0.0109)	0	-2.674^{***} (0.8096)
$\beta_{3,t}$ (curvature)	0	0	0.949^{***} (0.0238)	-1.758^{**} (1.3846)

(b) DNS-V variance matrix

	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$
$\beta_{1,t}$ (level)	0.0447^{***} (0.0151)	0	0
$\beta_{2,t}$ (slope)	0	0.0956^{***} (0.0216)	0
$\beta_{3,t}$ (curvature)	0	0	0.5250^{***} (0.0393)

NOTE: This table reports the estimates of the autoregressive model for the latent factors in the DNS-V model. Panel (a) shows the estimates for the state coefficient matrix Φ and the constant μ. Panel (b) shows estimates of the state variance matrix Σ_η. T-values are shown between brackets below the estimates. One asterisk (*) indicates a 90% significance level, two asterisks (**) indicate a 95% significance level and three asterisks (***) indicate a 99% significance level.

F Jacobian Matrix

DNS-L

For the DNS-L model the measurement equation is given by

$$Z_t(x_t) = \Lambda(\lambda_t)(\beta_{1,t}, \beta_{2,t}, \beta_{3,t})' = \Lambda_1(\lambda_t)\beta_{1,t} + \Lambda_2(\lambda_t)\beta_{2,t} + \Lambda_3(\lambda_t)\beta_{3,t}, \tag{63}$$

where $\Lambda_i(\lambda_t)$ is the i-th column of $\Lambda(\lambda_t)$.

$$\dot{Z}_t = \left.\frac{\partial Z(x_t)}{\partial x_t}\right|_{x=x_{t|t-1}} = \left[\left.\frac{\partial Z_t}{\partial \beta_{1,t}}\right|_{\beta_{1,t}=b_{1,t|t-1}} \quad \left.\frac{\partial Z_t}{\partial \beta_{2,t}}\right|_{\beta_{2,t}=b_{2,t|t-1}} \quad \left.\frac{\partial Z_t}{\partial \beta_{3,t}}\right|_{\beta_{3,t}=b_{3,t|t-1}} \quad \left.\frac{\partial Z(x_t)}{\partial \log(\lambda_t)}\right|_{\log(\lambda_t)=\log(\lambda_{t|t-1})} \right]$$

Now, the first three columns of $\dot{Z}_t$ follow straightforward from differentiating (65) to $\beta_{1,t}, \beta_{2,t}$ and $\beta_{3,t}$, respectively. For the fourth column we need to make use of the chain-rule in the following manner

$$\frac{\partial Z(x_t)}{\partial \lambda_t} = \frac{\partial Z(x_t)}{\partial \log(\lambda_t)} \cdot \frac{\partial \log(\lambda_t)}{\partial \lambda_t} = \frac{\partial Z(x_t)}{\partial \log(\lambda_t)} \cdot \frac{1}{\lambda_t}$$

$$\frac{\partial Z(x_t)}{\partial \log(\lambda_t)} = \frac{\partial Z(x_t)}{\partial \lambda_t} \cdot \lambda_t$$

Next, we apply again the chain rule and the quotient rule to obtain

$$\frac{\partial Z(x_t)}{\partial \lambda_t} = \frac{\beta_{2,t} \cdot \exp(-\tau \cdot \lambda_t)\,(\tau\lambda_t - \exp(\tau\lambda_t) + 1) + \beta_{3,t} \cdot \exp(-\tau\lambda_t)\left(\tau^2\lambda_t^2 + \tau\lambda - \exp(\tau\lambda_t) + 1\right)}{\tau\lambda_t^2}.$$

All together this yields the following result

$$\dot{Z}_t = \begin{bmatrix} \iota_{N\times 1} & \Lambda_2(\lambda_t) & \Lambda_3(\lambda_t) & \frac{\partial Z(x_t)}{\partial \lambda_t} \cdot \lambda_t \end{bmatrix}, \tag{64}$$

where $\iota_{N\times 1}$ is a vector of ones of length N.

DNS-LVM

For the DNS-LVM models the measurement equation is given by

$$Z_t(x_t) = \Lambda(\lambda_t)(\beta_{1,t}, \beta_{2,t}, \beta_{3,t})' + \Gamma\varepsilon_t^* = \Lambda_1(\lambda_t)\beta_{1,t} + \Lambda_2(\lambda_t)\beta_{2,t} + \Lambda_3(\lambda_t)\beta_{3,t} + \Gamma\varepsilon_t^*. \tag{65}$$

Compared to (65) this equation has the additional component $\Gamma\varepsilon_t^*$ and the state vector contains additional variables, such that $\dot{Z}_t$ gets additional elements. However, from the additional state variables, only ε_t^* is represented in the measurement equation. Hence only the derivative to this variable should be computed, all other additional elements of $\dot{Z}_t$ are equal to 0. Depending on the position in the state vector, the Jacobian matrix for the DNS-LVM(ES) model is

$$\dot{Z}_t = \begin{bmatrix} \iota_{N\times 1} & \Lambda_2(\lambda_t) & \Lambda_3(\lambda_t) & \frac{\partial Z(x_t)}{\partial \lambda_t} \cdot \lambda_t & \Gamma & 0_{N\times 5} \end{bmatrix}$$

and the Jacobian matrix for the DNS-LVM(FA) models is

$$\dot{Z}_t = \begin{bmatrix} \iota_{N\times 1} & \Lambda_2(\lambda_t) & \Lambda_3(\lambda_t) & \frac{\partial Z(x_t)}{\partial \lambda_t} \cdot \lambda_t & 0_{N\times 3} & \Gamma \end{bmatrix}$$