

Article

Recovering Yield Curves from Dynamic Term Structure Models with Time-Varying Factors

Hiroyuki Kawakatsu

Business School, Dublin City University, Dublin 9, Ireland; hiroyuki.kawakatsu@dcu.ie

Received: 29 July 2020; Accepted: 20 August 2020; Published: 22 August 2020



Abstract: A dynamic version of the Nelson-Siegel-Svensson term structure model with time-varying factors is considered for predicting out-of-sample maturity yields. Simple linear interpolation cannot be applied to recover yields at the very short- and long- end of the term structure where data are often missing. This motivates the use of dynamic parametric term structure models that exploit both time series and cross-sectional variation in yield data to predict missing data at the extreme ends of the term structure. Although the dynamic Nelson–Siegel–Svensson model is weakly identified when the two decay factors become close to each other, their predictions may be more accurate than those from more restricted models depending on data and maturity.

Keywords: yield curve; term structure; nonlinear filtering

1. Introduction

Yield curves need to be estimated, since bonds with different maturities are not directly comparable due to different coupon payments. Many central banks provide estimated yields from government bonds. The two common methods used to construct such estimates are those based on splines (smooth curve fitting) and those that are based on parametric models. Although most central banks make the estimated yield curve data publicly available, the exact methodology used to construct such estimates are usually not disclosed. For example, the U.S. Department of Treasury publishes daily treasury yield curve rates on their website (<https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield> (accessed on 14 June 2019)). The methodology used to obtain the data are explained at their Treasury Yield Curve Methodology page (<https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/yieldmethod.aspx> (last revised 14 October 2018; accessed on 14 July 2019)). which only states that the “yield curve is derived using a quasi-cubic hermite spline function” with no further details. Furthermore, “Treasury reserves the option to make changes to the yield curve as appropriate and in its sole discretion. Such changes may include, but are not necessarily limited to, adding, removing, or modifying inputs, and making changes to the methodology for deriving the yield curve.”

Many central banks make the estimated yield curve data publicly available. However, data are often missing, particularly at the very short and long end of the term structure. For example, for the (nominal) daily yield curve data from 3 January 2000 to 31 December 2018 (4753 observations) available from the U.S. Department of Treasury, there are 396 (8%) missing values for the one-month rate and 994 (21%) missing values for the 30-year rate. This is unfortunate as the rates at these very short and long ends of the term structure are often of interest.

Given the black box nature of yield curve data construction, the main objective of this paper is to examine whether parametric dynamic term structure models can recover the yield curve data provided by central banks. As the yield curve data are estimated from the traded bond prices, one would ideally try to recover the yield estimates from market quotes of the traded bond prices. However, such data are usually proprietary and not easily accessible to academic researchers. Furthermore, as emphasized

in Nymand-Andersen [1] “data cleaning is too often neglected or deprioritised [...] with the risk of not being able to differentiate noise from real (economic) signals when interpreting statistical results”.

The approach taken in this paper is to use the publicly provided data to infer the full term structure of yields by predicting the missing values. As yield curve data have a panel data structure, we can exploit both the time series and cross-sectional variation in doing so. Recent work by Koo et al. [2] do this nonparametrically without imposing parametric functional forms. However, they assume that time variation in yields is driven by a vector of common observed covariates (including deterministic trends). They apply their method to the daily bond quotes data from the proprietary CRSP U.S. Treasury database.

This paper takes a parametric approach and considers some generalization of the class of dynamic term structure models used in the literature. Section 2 summarizes this literature that this paper builds upon. The main novelty of the specification considered in this paper is to allow the two decay parameters in the generalized Nelson and Siegel [3] model of Svensson [4], Söderlind and Svensson [5] to be time-varying. Koopman et al. [6] considered a specification with one time-varying decay parameter that generalizes the Nelson and Siegel [3] model. The two decay parameter model is harder to estimate than the one decay parameter model due to weak identification when the two decay parameters take similar values, as discussed in Section 2.

Models with time-varying decay parameters become nonlinear in the state variables of a state-space model. The issue of nonlinear filtering and estimation is discussed in Section 3. Section 4 conducts a small Monte Carlo simulation experiment to compare the finite sample performance of alternative nonlinear filters. An important criterion in choosing among the alternative filters is computational cost (efficiency), not just their accuracy. The results of the simulations indicate that the extended Kalman filter can be an effective choice due its computational simplicity.

An extensive literature compares the performance of various term structure models using time series out-of-sample prediction for future observations as performance criterion [1,7]. This papers uses cross-sectional out-of-sample prediction for yields with maturity not used in estimation (but in the same sample period) as performance criterion.

Section 5 compares the out-of-sample cross-sectional predictability of alternative dynamic term structure specifications while using two types of data. For data known to be generated from parametric models, the two time-varying decay parameter specification performs best, despite the weak identification problem. This result is perhaps not surprising. The results using data estimated from spline based methods are more mixed and can be data dependent. If the out-of-sample maturity is inside the range of maturities used for estimation, a simple linear interpolation may perform as well as predictions from parametric dynamic term structure models. However, interpolation cannot be used for maturities at the very short- or long- end of the term structure where the missing data problem is most prevalent. For these extreme maturities, predictions from the general specification with two time-varying decay parameters generally perform better than those based on more restricted models. However, this is not always the case and the simpler models can perform as well as the generalized model that is proposed in this paper.

2. Yield Curve Estimation Methods

There are two main approaches to yield curve modeling: spline based and parametric model based [1]. The spline based approach is used by the Bank of England, the Federal Reserve Bank of New York, Bank of Canada, among others (footnote 2, p. 5 in [1]). The parametric model based approach is used by many European central banks, such as Deutsche Bundesbank, Banco de España, Banca d'Italia, Banque de France, and the European Central Bank (footnote 3, p. 5 in [1]).

2.1. Spline Based Methods

The spline based approach uses piecewise polynomials to flexibly fit a curve through the data. The splines ensure that the piecewise functions are smoothly joined at the knot points, resulting in

a smooth curve over the support. To implement this method, one needs to specify the number of knot points and their positions. The piecewise polynomials may be fit to the forward rate function, the discount function, or the log of the discount function [8].

The spline based methods differ in how they determine the coefficients for each piecewise polynomials. The different methods can be characterized by the use of different regularization, or roughness penalty. The roughness penalty function may or may not depend on maturity or vary over time (Section 3.1 in [1]). Related to spline based methods are the nonparametric methods, where polynomials are locally weighted by kernel functions [2,9]. The analogue of roughness penalty for nonparametric methods is the bandwidth, which controls the ‘locality’ of the kernel weights.

The flexibility to fit any curve is both an advantage and disadvantage of the spline based approach. It is well known that the shape of the fitted curve can be quite sensitive to the choice of roughness penalty. “Given the superior curvature properties of cubic splines it is a bit surprising how relatively poor the spline smoother performs [in the simulations]. [...] One explanation of these results is therefore that this is related to the lack of a good bandwidth estimator”. ([9], p. 215).

For researchers to be able to reproduce the yield curve data obtained from spline based methods, we need information on the position of the knot points, type of spline (piecewise polynomial) used, and the weight used for the roughness penalty term. I am not aware of any central bank that uses spline based methods that provides such reproducible information. This is in contrast to the transparency of the parsimonious parametric approach that is discussed below.

2.2. Parametric Approach

The most commonly used parametric models for the yield curve are the Nelson and Siegel [3] and Svensson [4], Söderlind and Svensson [5] models. One of the appealing features of these models is that one can recover the implied instantaneous forward rate from the yield curve and vice versa. The Nelson and Siegel [3] instantaneous forward rate at time t with maturity τ is specified as

$$f_t(\tau) = \beta_{1t} + \beta_{2t}e^{-\lambda_t\tau} + \beta_{3t}\lambda_t\tau e^{-\lambda_t\tau}$$

The four parameters, known as latent factors, β_{1t} , β_{2t} , β_{3t} , λ_t have a t subscript to indicate that they can be time varying.

The yield curve at t with maturity τ is then the ‘average’ of forward rates given by

$$\begin{aligned} y_t(\tau) &= \frac{1}{\tau} \int_0^\tau f_t(u) du \\ &= \beta_{1t} + \beta_{2t} \frac{1}{\lambda_t \tau} (1 - e^{-\lambda_t \tau}) + \beta_{3t} \frac{1}{\lambda_t \tau} (1 - e^{-\lambda_t \tau} - \lambda_t \tau e^{-\lambda_t \tau}) \end{aligned}$$

The yield curve is linear in the factors β_{1t} , β_{2t} , β_{3t} but nonlinear in the decay factor λ_t . Associated with each factor β are the ‘loadings’

$$\begin{aligned} w_{1t}(\tau) &= 1 \\ w_{2t}(\tau) &= \frac{1}{\lambda_t \tau} (1 - e^{-\lambda_t \tau}) \\ w_{3t}(\tau) &= \frac{1}{\lambda_t \tau} (1 - e^{-\lambda_t \tau} - \lambda_t \tau e^{-\lambda_t \tau}) = w_{2t}(\tau) - e^{-\lambda_t \tau} \end{aligned}$$

Not only does this parametric form turn out to fit the yield data flexibly, but the factors also have natural interpretations. This interpretability has contributed to the wide spread use of the Nelson and Siegel [3] model. The decay factor λ_t controls how the yield curve exponentially decays with maturity τ . β_{1t} can be interpreted as the level, or long term, factor since an increase in β_{1t} will shift the yield for all maturities uniformly due to the constant loading $w_{1t} = 1$. β_{2t} is known as the slope, or short term, factor. w_{2t} is an exponentially decreasing function of τ with $w_{2t}(0) = 1$ and

$w_{2t}(\infty) = 0$. An increase in β_{2t} shifts the short rate more than the long rate, and hence changes the slope of the yield curve. β_{3t} is the curvature, or medium term, factor. $w_{3t}(0) = w_{3t}(\infty) = 0$ and $w_{3t}(\tau)$ is concave with a local maximum (hump shape). The decay rate λ_t determines the position of the local maximum. An increase in β_{3t} shifts the medium rates more than the short and long rates and, hence, increases the curvature of the yield curve.

These parametric loadings are known to match the data quite well. The first three principal components of the sample covariance of yields account for most of the variation in the U.S. yields, as documented by Litterman and Scheinkman [10]. The first factor is a ‘parallel shift’ and accounts for 80–90% of the term structure variation. The second factor is a ‘twist’ where the short and long rates move in opposite directions and accounts for 5–10% of variation. The third factor is a ‘butterfly’ where intermediate rates move in the opposite direction of the short and long rates and accounts for 1–2% of the variation.

The Svensson [4], Söderlind and Svensson [5] model adds another factor β_{4t} and associated decay parameter λ_{2t} with forward rate specified as

$$f_t(\tau) = \beta_{1t} + \beta_{2t}e^{-\lambda_{1t}\tau} + \beta_{3t}\lambda_{1t}\tau e^{-\lambda_{1t}\tau} + \beta_{4t}\lambda_{2t}\tau e^{-\lambda_{2t}\tau}$$

and yield curve

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\frac{1}{\lambda_{1t}\tau}(1 - e^{-\lambda_{1t}\tau}) + \beta_{3t}\frac{1}{\lambda_{1t}\tau}(1 - e^{-\lambda_{1t}\tau} - \lambda_{1t}\tau e^{-\lambda_{1t}\tau}) + \beta_{4t}\frac{1}{\lambda_{2t}\tau}(1 - e^{-\lambda_{2t}\tau} - \lambda_{2t}\tau e^{-\lambda_{2t}\tau}) \quad (1)$$

The Nelson and Siegel [3] yield curve can have, at most, one hump. The additional two factors in the Svensson [4], Söderlind and Svensson [5] model allows the yield curve to have more than one hump.

Although the additional factors increase the flexibility of the model to fit the data, they make the model potentially weakly identified. The loadings

$$w_{3t}(\tau) = \frac{1}{\lambda_{1t}\tau}(1 - e^{-\lambda_{1t}\tau} - \lambda_{1t}\tau e^{-\lambda_{1t}\tau})$$

$$w_{4t}(\tau) = \frac{1}{\lambda_{2t}\tau}(1 - e^{-\lambda_{2t}\tau} - \lambda_{2t}\tau e^{-\lambda_{2t}\tau})$$

associated with the factors β_{3t} , β_{4t} become highly collinear for $\lambda_{1t} \approx \lambda_{2t}$ [11]. Despite this weak identification problem, most central banks that adopt the model based approach appear to favor the extended Svensson [4], Söderlind and Svensson [5] model over the original Nelson and Siegel [3] model [1].

Another recent contributing factor for the popularity of this class of parametric models is the relation with no arbitrage restrictions. Early work conducted by Filipovic [12] showed that, if the latent factors follow diffusion processes, the Nelson and Siegel [3] model is generally not arbitrage free. To be consistent with no arbitrage, the factors need to be constant or deterministic. For the Svensson [4], Söderlind and Svensson [5] model, only one factor can be non-deterministic to be consistent with no arbitrage. Christensen et al. [13] are able to find a member of the affine class of arbitrage free models of Duffie and Kan [14] with factor loadings that exactly match those of the Nelson and Siegel [3] model. However, this requires modifying the Nelson and Siegel [3] model by including an additional deterministic term in the yield curve that only depends on maturity. Krippner [15] shows how the Nelson and Siegel [3] factors (approximately) map to those of the arbitrage free gaussian affine class models of Dai and Singleton [16].

For the Svensson [4], Söderlind and Svensson [5] model, both Christensen et al. [17] and Krippner [15] show that, in addition to the maturity depend yield adjustment term, a second

slope factor is needed to be consistent with no arbitrage. In these so-called arbitrage free affine dynamic Nelson–Siegel–Svensson models, the factors follow diffusion processes, except for the decay parameters, which remain fixed.

These parametric models can be fit to a cross-section of bonds period by period. A dynamic term structure model links these factors over time by assuming that they follow an autoregressive process. Koopman et al. [6] considered a dynamic version of the Nelson and Siegel [3] model where all latent factors follow an AR(1) process. Christensen et al. [17] considered dynamic versions of the Svensson [4], Söderlind and Svensson [5] model. However, they restricted the two decay factors to be fixed parameters. This paper considers a model where the two decay factors also follow an AR(1) process.

Early work by Diebold and Li [18] used a two-step method to estimate dynamic term structure models. The two-step method exploits the fact that, conditional on the decay factor, the model becomes linear in the remaining factors. However, in the first step, one needs to arbitrarily fix the value of the decay parameter to some ‘reasonable’ value. Swanson and Xiong [19] use a three-step procedure, in which the decay parameter for the two-step procedure is chosen using a grid search that optimizes the in-sample fit of the factors. This appears to be done period by period recursively without specifying the dynamics of the decay parameters. Statistical inference based on these multistep procedures is complicated, since the estimates in each step are conditional on the estimates from the previous step(s).

More recent work jointly estimate all parameters by maximizing a gaussian likelihood function [6,13,17]. The likelihood function is evaluated by casting the dynamic term structure model in state-space form. For the fixed decay factor specifications of Christensen et al. [13,17], the standard linear Kalman filter can be used to evaluate the likelihood. For specifications with time-varying decay factors, the state-space model becomes nonlinear in the states and the standard Kalman filter cannot be applied. Koopman et al. [6] use the extended Kalman filter that applies the Kalman filter to a linearized approximation.

One main reason Christensen et al. [13,17] to restrict the decay parameter(s) to be fixed is to exploit the no arbitrage condition. They show that their specifications belong to the class of affine arbitrage free models provided a yield adjustment term is added to the original Nelson–Siegel–Svensson specifications. There is some evidence that this yield adjustment term may be economically small and, hence, the gains from exploiting the no-arbitrage restriction may also be small [20]. This paper follows the lead of Koopman et al. [6] and examines the gain, if any, of allowing the decay parameters to be time varying instead.

An important aspect of the model based approach not emphasized enough in the literature is its parsimony. In particular, for researchers to be able to reproduce the yields and forward rates for any maturity, all one needs are the values of a small number of latent factors (six for the Svensson [4], Söderlind and Svensson [5] model). The European Central Bank is commended to not only make the daily yield curve data publicly available, but to also provide the daily estimated values of the six latent factors of the Svensson [4], Söderlind and Svensson [5] model used to fit the yield curve (<http://sdw.ecb.europa.eu/browse.do?node=9691126>).

2.3. Which Method to Use?

As evidenced by the fact that both spline based and parametric model based approaches are used in practice, neither approach dominates the other. The spline based approach is favored when the emphasis is on obtaining a ‘good’ fit that passes through most of the data points. The model based approach is favored when the emphasis is on parsimony and smoothness of the resulting curve avoiding overfitting in-sample. The appropriate choice requires trading-off these pros and cons, depending on one’s primary objective for fitting the yield curve [1]. Alternatively, rather than choosing one method over the other, one can combine or average the estimates from both methods [19].

3. Dynamic Term Structure Models with Time-Varying Factors

3.1. State-Space Formulations

I compare the performance of four parametric models with time-varying factors to examine the cross-sectional predictability of dynamic term structure models. All of these models can be written in state-space form, as follows.

$$y_t = Z(\alpha_t) + \epsilon_t, \quad \epsilon_t \sim N(0, H) \quad (2)$$

$$\alpha_{t+1} = d + T\alpha_t + v_t, \quad v_t \sim N(0, Q), \quad v_t \perp \epsilon_t \quad (3)$$

for $t = 1, \dots, n$. y_t is the $m \times 1$ observed vector of yields in period t with maturities $\tau = (\tau_1, \dots, \tau_m)^\top$ and $\alpha_t = (L_t, S_t, C_{1t}, C_{2t}, \log \lambda_{1t}, \log \lambda_{2t})$ is the 6×1 unobserved latent state vector. The components of the state vector are the level (L_t), slope (S_t), curvature (C_{1t}, C_{2t}), and decay ($\lambda_{1t}, \lambda_{2t}$) factors. $v_t \perp \epsilon_t$ denotes statistical independence of v_t and ϵ_t . The parameters of the model are $\theta = (H, d, T, Q)$. The $m \times m$ symmetric positive definite matrix H is assumed to be diagonal with positive entries and Q is a 6×6 symmetric semi-positive definite matrix. If we do not restrict T or Q to be diagonal, there are $m + 6 + 6 \times 6 + 6 \times 7/2 = m + 63$ parameters to estimate.

The measurement equation links the yields to the state (factor) vector via the loadings

$$\begin{aligned} Z(\alpha_t) &= 1_m L_t + f(\lambda_{1t}, \tau) S_t + g(\lambda_{1t}, \tau) C_{1t} + g(\lambda_{2t}, \tau) C_{2t} \\ f(\lambda, \tau) &= \frac{1}{\lambda \tau} (1 - e^{-\lambda \tau}) \\ g(\lambda, \tau) &= f(\lambda, \tau) - e^{-\lambda \tau} \end{aligned}$$

where 1_m is the $m \times 1$ vector of ones.

The models used in this paper do not impose the no arbitrage restrictions. As shown in Christensen et al. [17], the model with two decay factors λ_1, λ_2 but one slope factor S_t cannot be arbitrage free. Christensen et al. [17] introduce a second slope factor in addition to the yield adjustment term to make the model arbitrage free. However, the model with a single slope factor already has a large number of parameters ($m + 63$) to estimate. Adding a second slope factor will increase the number of parameters to $m + 7 + 49 + 28 = m + 84$ (without restricting T or Q). Another reason to keep the model simple is to avoid exacerbating the weak identification problem alluded to in Section 2. The arbitrage free model with two decay parameters and two slope factors is not identified without imposing an identification restriction, such as $\lambda_1 > \lambda_2$. Furthermore, there is some evidence at least with U.S. data that the Nelson and Siegel [3] model is ‘nearly’ arbitrage free indicating that the yield adjustment term may be economically small [20]. For these reasons, I keep the original Nelson–Siegel–Svensson specification without imposing the no arbitrage restrictions. As discussed below, I instead examine the gains from having a time-varying decay parameter.

An issue that does not appear to be discussed much in the literature is whether the specification should restrict the yield to be non-negative. Most dynamic term structure specifications appear not to impose such non-negativity restriction. Inspecting the factor data publicly available from the European Central Bank indicates that the level factor L_t and decay parameters λ_1, λ_2 are restricted to be non-negative (though I could not find any documentation that mentions this restriction). Koopman et al. [6] do not appear to restrict either the slope factor L_t or the decay factor λ_t to be non-negative. The specifications used in this paper do not restrict the level factor L_t but do restrict the decay parameters λ to be non-negative. This is done by entering the decay parameters in the state vector α_t with the (natural) log transformation $\log \lambda$.

The dynamic Nelson–Siegel–Svensson (DNSS) specification presented above is the most general model considered in this paper. I also consider three special cases of this general model that have been used in the literature. The Koopman et al. [6] specification nulls out rows and columns four and six of the parameters d, T , and Q resulting in a state-space model with a four dimensional state vector

$\alpha_t = (L_t, S_t, C_{1t}, \log \lambda_{1t})$. This specification is called the dynamic Nelson–Siegel model, denoted *DNS*, and differs from that in Koopman et al. [6] in that the decay factor λ_t is restricted to be non-negative. The two other special cases keep the decay parameters λ fixed as in Christensen et al. [13,17], but without imposing the no arbitrage restrictions. These specifications are referred to as the Nelson–Siegel–Svensson (*NSS*) model with state vector $\alpha_t = (L_t, S_t, C_{1t}, C_{2t})$ and the Nelson–Siegel (*NS*) model with state vector $\alpha_t = (L_t, S_t, C_{1t})$.

3.2. Filtering and Estimation

The parameters of the state-space model (2) and (3) are estimated by maximizing the gaussian log-likelihood function. For models with time-varying decay factors, the measurement equation is not linear in the state vector. The extended Kalman filter approach takes a linear approximation of measurement equation about the predicted state vector $a_{t|t-1} = E_{t-1}[\alpha_t]$ as

$$\begin{aligned} y_t &= Z(\alpha_t) + \epsilon_t \\ &\approx Z(a_{t|t-1}) + \frac{\partial Z}{\partial \alpha^\top} (\alpha_t - a_{t|t-1}) + \epsilon_t \\ &= c_t + \frac{\partial Z}{\partial \alpha^\top} \alpha_t + \epsilon_t, \quad c_t \equiv Z(a_{t|t-1}) - \frac{\partial Z}{\partial \alpha^\top} a_{t|t-1} \end{aligned}$$

where the $m \times 6$ jacobian $\partial Z / \partial \alpha^\top$ is evaluated at $\alpha = a_{t|t-1}$. The expressions for the jacobian $\partial Z / \partial \alpha^\top$ are given in online Appendix A.1.

The gaussian log-likelihood function can be evaluated using the extended Kalman filter, which consists of the filtering steps

$$\begin{aligned} e_t &= y_t - c_t - \frac{\partial Z}{\partial \alpha^\top} a_{t|t-1} = y_t - Z(a_{t|t-1}) \\ F_t &= \frac{\partial Z}{\partial \alpha^\top} A_{t|t-1} \frac{\partial Z}{\partial \alpha} + H \\ a_{t|t} &= a_{t|t-1} + A_{t|t-1} \frac{\partial Z}{\partial \alpha^\top} F_t^{-1} e_t \\ A_{t|t} &= A_{t|t-1} - A_{t|t-1} \frac{\partial Z}{\partial \alpha^\top} F_t^{-1} \frac{\partial Z}{\partial \alpha} A_{t|t-1} \end{aligned}$$

and prediction steps

$$\begin{aligned} a_{t+1|t} &= d + T a_{t|t} \\ A_{t+1|t} &= T A_{t|t} T^\top + Q \end{aligned}$$

The filter is initialized with $a_{1|0} = (I - T)^{-1}d$ and the solution to the Lyapunov equation $A_{1|0} = T A_{1|0} T^\top + Q$. The log-likelihood is evaluated as

$$\ell = -3n \log(2\pi) - \frac{1}{2} \sum_{t=1}^n (\log |F_t| + e_t^\top F_t^{-1} e_t) \quad (4)$$

The affine arbitrage free models estimated in Christensen et al. [13,17] have the important feature that the decay parameters λ_1, λ_2 are fixed and not time-varying. This makes the state-space model linear in the states (though still nonlinear in the parameters) and the standard Kalman filter can be used to evaluate the likelihood function. Koopman et al. [6] consider the Nelson and Siegel [3] model with time-varying decay parameter λ_t and find improvement in the in-sample fit of the model. Once we make the decay parameter part of the state vector α_t , the state-space model becomes nonlinear in the state and the standard Kalman filter cannot be used. Koopman et al. [6] use the extended Kalman filter to the linearized model.

The extended Kalman filter is computationally attractive, since it applies the standard Kalman filter to the approximated linear state-space model. However, the concern is the error due to the linear approximation. As an important feature of the Nelson–Siegel–Svensson specification is the ability to parsimoniously capture nonlinear curvatures in the term structure, one may be concerned whether the linearization in the extended Kalman filter may compromise this feature. As this issue appears to be little explored in the literature, I consider two alternative nonlinear filtering methods [21,22]: the unscented filter and the Rao-Blackwellized particle filter.

The unscented Kalman filter approximates the unknown distribution of a nonlinearly transformed random variable by a small number of discretized points and associated weights, known as sigma points and sigma weights. These points and weights are chosen to match the first two moments of the target distribution. It can be shown that the mean of the extended Kalman filter estimate is accurate to first order, while that of the unscented filter is accurate to the second order. (The variance is estimated to the second order of approximation by both filters.) Further details are provided in Appendix A.2.

To apply the unscented filter, one needs to specify the sigma points and weights. The requirement to match the first two moments of the target distribution imposes certain restrictions on the choice. However, they do not uniquely narrow down the choice and certain ‘tuning’ parameters need to be selected by the user. One of the simplest and commonly used choice suggested in Julier and Uhlmann [23] requires choosing how to allocate the non-negative weights that sum to one over the sigma points. The first sigma point is equal to the mean of the target distribution while the remaining number of points come in pairs that offset each other, so that the mean of the sigma points is equal to the target mean.

The second alternative is based on particle filtering, also known as sequential Monte Carlo [24,25]. The unscented filter uses a small number of fixed points and weights to approximate the unknown distribution. In contrast, particle filters use a large number of random draws to accurately approximate the unknown target distribution. In a nonlinear state-space model, the target distribution of interest is the distribution of the (unobserved) state vector α_t . In a standard application of particle filters, one would draw particles from the full state space with dimension equal to that of α_t .

Rao–Blackwellization draws particles from a reduced dimension by partitioning the state space [26]. By drawing a subset of state variables conditional on the other state variables, one can reduce simulation variance and obtain much more accurate filtered estimates than drawing from the full state space without conditioning. The key to achieving effective variance reduction through Rao–Blackwellization is to find an appropriate partitioned state space. An important feature of the DNSS model is that conditional on the two decay factors $\lambda_t = (\lambda_{1t}, \lambda_{2t})$, the model becomes linear in the remaining four state variables L_t, S_t, C_{1t}, C_{2t} . To denote the partitioned system, split the state vector as $\alpha_t = (x_t, \lambda_t)^\top$ where $x_t = (L_t, S_t, C_{1t}, C_{2t})$ and $\lambda = (\log \lambda_{1t}, \log \lambda_{2t})$. The system can then be written as

$$\begin{aligned} y_t | \alpha_t &\sim N(Z(\lambda_t)x_t, H) \\ x_t | \lambda_t, \alpha_{t-1} &\sim N(\mu_{1t} + Q_{12}Q_{22}^{-1}(\lambda_t - \mu_{2t}), Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}) \\ \lambda_t | \alpha_{t-1} &\sim N(\mu_{2t}, Q_{22}) \end{aligned}$$

where $Z(\lambda_t) = (1_m, f(\lambda_{1t}, \tau), g(\lambda_{1t}, \tau), g(\lambda_{2t}, \tau))$ is $m \times 4$ and $\mu_t = d + T\alpha_{t-1} = (\mu_{1t}, \mu_{2t})$. The conditional mean of $x_t | \lambda_t, \alpha_{t-1}$ can be written as

$$\begin{aligned} E[x_t | \lambda_t, \alpha_{t-1}] &= \mu_{1t} + Q_{12}Q_{22}^{-1}(\lambda_t - \mu_{2t}) = d_x(\lambda_t, \lambda_{t-1}) + T_x x_{t-1} \\ d_x(\lambda_t, \lambda_{t-1}) &= d_1 + T_{12}\lambda_{t-1} + Q_{12}Q_{22}^{-1}(\lambda_t - d_2 - T_{22}\lambda_{t-1}) \\ T_x &= T_{11} - Q_{12}Q_{22}^{-1}T_{21} \end{aligned}$$

Details of the filtering algorithm, mostly taken from Schon et al. [26], are reproduced in Appendix A.3 for the reader's convenience. The extended and the unscented Kalman filters can both be considered to be Rao–Blackwellized for the *DNSS* model as the state equation is linear in the states. The approximation is only applied to the measurement equation, which is nonlinear in the state variable(s).

3.3. Computational Issues

As already indicated, the most general *DNSS* specification has a large number of parameters ($m + 63$) to estimate. In order to keep the number of parameters linear in the state dimension for estimation purposes, I restrict T and Q in the state equation to be both diagonal as in Christensen et al. [17]. This reduces the number of parameters to estimate for the *DNSS* model to the more manageable $m + 18$.

Even with this restricted specification, the *DNSS* model with two time-varying decay factors is significantly harder to estimate than the *DNS* model (with one time-varying decay factor). The reason is the weak identification problem with the *DNSS* model when $\lambda_{1t} \approx \lambda_{2t}$. The consequence is that the numerical optimization procedure can fail to converge when the starting parameter values result in $\lambda_{1t} \approx \lambda_{2t}$. As a partial solution to this problem, optimization of the log-likelihood is carried out for a number of randomly drawn sets of starting values.

The three-step estimation procedure in Swanson and Xiong [19] deals with this issue by restricting the two decay parameters to be a certain distance apart period-by-period ([19], footnote 8, p. 12). However, when the dynamics of λ_t are specified as an autoregressive processes, as is done in this paper, such period-by-period restriction cannot be imposed. Therefore, I effectively do a grid search over a large number of randomly drawn starting parameter values.

4. Simulation Study

4.1. Comparison of Nonlinear Filters

This section conducts a small Monte Carlo simulation study to compare the accuracy of the alternative nonlinear filtering methods described in the previous section. Due to computational cost, this simulation study is restricted to studying the accuracy of the filtering algorithms assuming the true parameter values are known. In practice the parameter values are unknown and must be estimated. The effect of filtering error on the accuracy of the parameter estimates is an important issue to be examined, but is left for future work. The main computational difficulty for examining estimation error is the Rao–Blackwellized particle filter as it requires applying the recursive filter from drawing a large number of particles for each simulated sample, a simulation within a simulation problem. As shown below, even with known parameter values, the Rao–Blackwellized particle filter is considerably more computationally expensive than the other nonlinear filters.

4.2. Simulation Design

In the simulation experiment, I consider two weighting allocations for the unscented filter. The first is to equally weight all points and it is denoted *USF*. The second is to allocate half weight to the first sigma point (which is the target mean) and distribute the remaining half equally to the remaining points. This second filter is denoted *USF*($w_0 = 0.5$).

The following simulation design was used for drawing simulated sample observations. For each replication, a sample of size $n = 3000$ was simulated from the *DNSS* model with parameter values $H = 10^{-5}I_m$, $Q = 10^{-2}I_6$, $T = 0.9I_6$, $d = (I_6 - T)E[\alpha_t]$,

$$E[\alpha_t] = (2.6, -1.8, 11.7, -11.3, 0.97, 0.83)$$

where I_m is the $m \times m$ identity matrix. The values of the state variable (unconditional) means were taken from the sample (6 September 2004 to 21 February 2019) means of the daily factor data from the European Central Bank.

From the state variables, two sets of yield curves were simulated one with nine maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ and one with 13 maturities $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$ both in years. Using the simulated yields y_t and the ‘true’ parameter values $\theta = (H, d, T, Q)$ the filtered states $a_{t|t}$ were obtained using the extended Kalman filter (denoted *EKF*), unscented filter (*USF*, $USF(w_0 = 0.5)$), and the Rao–Blackwellized particle filters (denoted *RBP*(m), where m is the number of particles used).

To make the accuracy measures comparable across different state variables, I use the scale-free accuracy measure recommended by Hyndman and Koehler [27]. Let $e_t = a_{t|t} - \alpha_t$ denote the error of the filtered state series $a_{t|t}$. The scaled error is defined as $q_t = e_t / mae_{rw}$, where $mae_{rw} = \sum_{t=2}^n |\alpha_t - \alpha_{t-1}| / (n - 1)$ is the mean absolute error from a ‘naive’ forecast assuming α_t follows a random walk. The scale-free accuracy measure used is the mean absolute scaled error defined as $mase = \sum_{t=1}^n |q_t| / n$.

A sample of size $n = 3000$ is simulated 999 times and, for each sample, the *mase* is computed using the alternative nonlinear filters to obtain a distribution of *mase* for each nonlinear filter. Figure 1 compares the distribution of *mase* for the measurement with nine maturities and Figure 2 for the measurement with 13 maturities. Several important features are apparent in these figures. First, no filter dominates in accuracy for all state variables. For recovering the (conditionally) linear state variables corresponding to the L , S , C_1 , C_2 factors, the approximate filters *EKF*, *USF* are competitive against *RBP*($m = 999$). However, for the decay factors λ , which are the main sources of nonlinearity, *RBP*($m = 999$) is more accurate than the approximate filters *EKF*, *USF*.

The *mase* accuracy measure uses the random walk forecast as benchmark and errors with $mase < 1$ indicate better recovery than from the random walk model. The figures show that this condition is clearly satisfied only for the level factor L (the state variables were generated from a persistent but stationary AR(1) model with $T = 0.9$). The potentially collinear two curvature factors C_1 , C_2 have *mase* of at least two confirming the weak identification problem with the *DNSS* model.

Finally, the performance of the filters depend on the signal to noise ratio in the observed measurements. This signal-to-noise ratio depends on the measurement error variance H . The value $H = 10^{-5}$ used in the figures was chosen, so that some differences among the filters become visually apparent. The distribution of filtering errors for lower signal-to-noise ratio with $H = 0.1$ (not reported) are visually indistinguishable and overlap each other.

4.3. Trading off Accuracy and Computational Cost

The main finding from this small simulation experiment is that no filter dominates in terms of accuracy ranking. Another important consideration in choosing the appropriate filter is its computational cost. Some filters, like *EKF*, have no tuning parameter to control its accuracy (except possibly to use higher order expansions) and its performance depends on the extent of nonlinearity in the data. For other filters, we have some control over their accuracy by the choice of tuning parameters. For example, for *USF* the number of sigma points, their positions, and associated weights can affect their accuracy. Generally, there would be a tradeoff between accuracy and computational cost. For *USF*, the more points you use, we expect better accuracy, but at a higher computational cost (longer computing time).

Figures 3 and 4 compare the accuracy versus computational cost tradeoff of the nonlinear filters. Because the Rao–Blackwellized filter (*RBP*) takes much more longer than the other filters, the computation time on the horizontal axis is the log of timing relative to the extended Kalman filter. A fast and accurate filter would appear in the bottom left corner of each plot. Filters that lie to the north-west (up-right) of a filter are dominated by that filter.

Viewed from this accuracy-computational cost tradeoff criterion, Figures 3 and 4 show that *RBP* does not dominate any other filter, but it is dominated for several state variables for the *DNSS* model. In particular, the *RBP* filter with $m = 99$ particles is always dominated, except for the λ_2 state variable with nine yields. The computational cost of the *RBP* filter is roughly proportional to $2nm$, where n is the sample size, two is the dimension of the simulated state vector (the two decay parameters), and m

the number of particles. *RBP* requires running the filter for each particle over the sample n and, hence, the computational cost of order $O(2\text{ nm})$. For the daily sample analyzed below, $n \approx 3000$ as used in the simulations. Although computation time is highly dependent on software and hardware used to implement the filters, in my implementation filtering 13 maturities of sample size $n = 3000$ takes about 0.07 seconds for the *EKF* filter, 0.16 s for the *USF* filter (13 sigma points), and 27.5 s for *RBP* ($m = 999$). *RBP* ($m = 999$) is about 400 times slower than *EKF*.

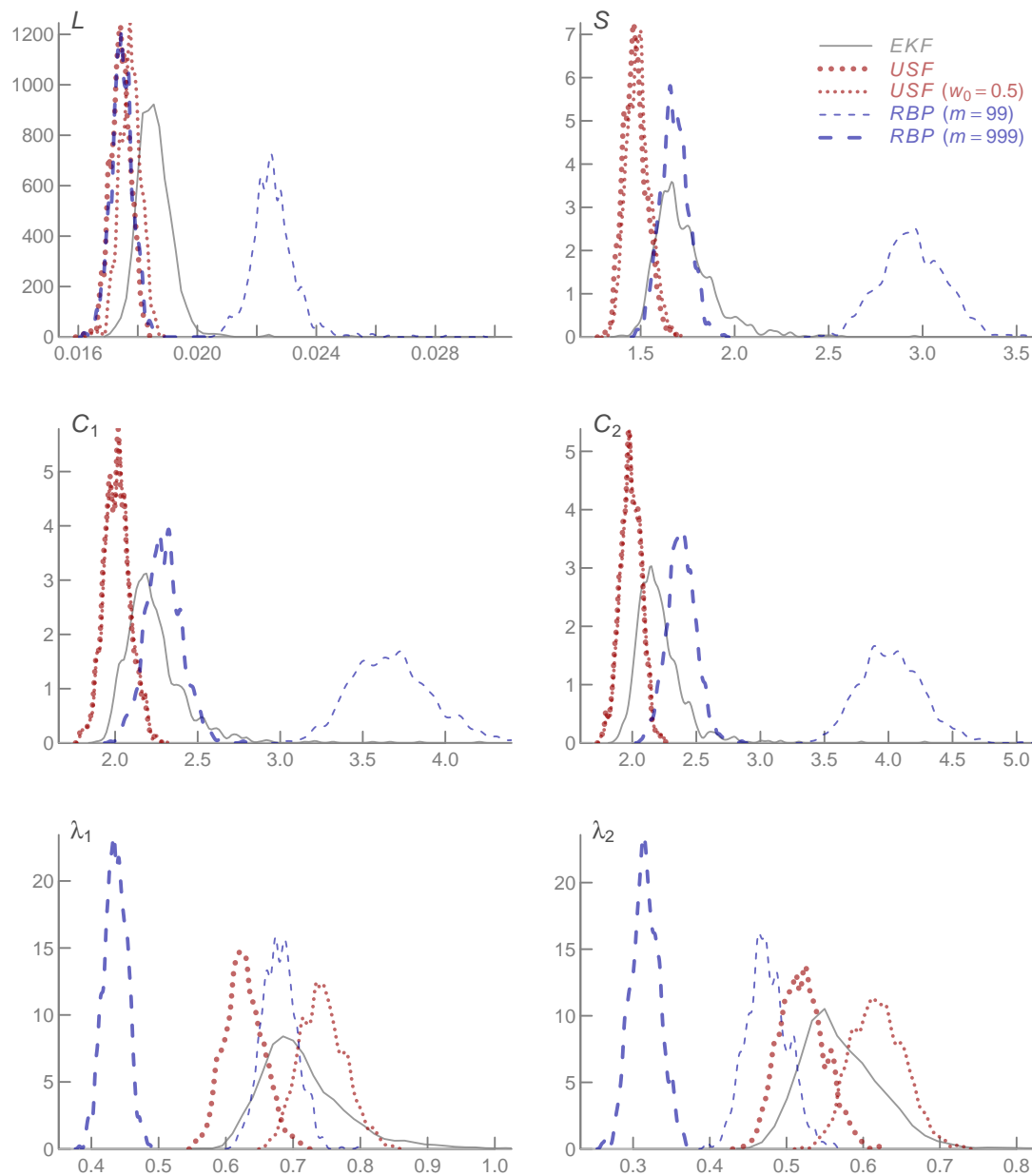


Figure 1. Distribution of mean absolute scaled errors of filtered state series over 999 simulated draws from the Nelson–Siegel–Svensson dynamic factor model. The measurement equation consists of 9 yields with maturities $\tau = 1/12, 6/12, 1, 2, 3, 5, 7, 10, 20$ years each with 3000 observations. L is the level, S the slope, C_1 , C_2 curvature factors with decay parameters λ_1 , λ_2 , respectively. Each panel compares the error distribution of the filtered series over five different filters. *EKF* is the extended Kalman filter, *USF* is the unscented filter with equal weights on every sigma point, *USF* ($w_0 = 0.5$) is the unscented filter with weight w_0 on the central (first) sigma point and equal weights for the remaining sigma points, *RBP* is the Rao–Blackwellized particle filter with m particles.

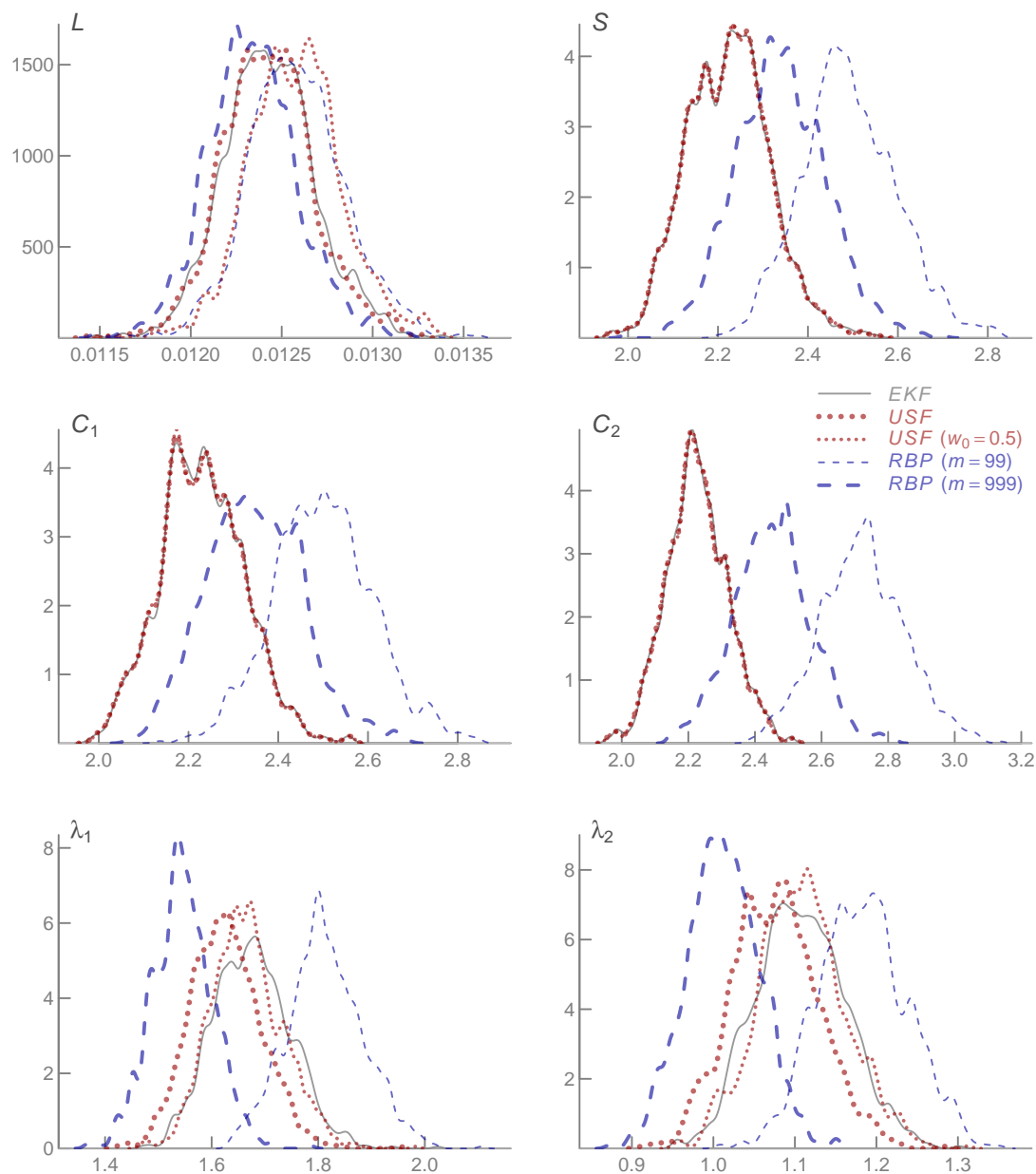


Figure 2. Distribution of mean absolute scaled errors of filtered state series over 999 simulated draws from the Nelson-Siegel-Svensson dynamic factor model. The measurement equation consists of 13 yields with maturities $\tau = 1, 3/2, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25$ years each with 3000 observations. L is the level, S the slope, C_1 , C_2 curvature factors with decay parameters λ_1 , λ_2 , respectively. Each panel compares the error distribution of the filtered series over five different filters. *EKF* is the extended Kalman filter, *USF* is the unscented filter with equal weights on every sigma point, *USF* ($w_0 = 0.5$) is the unscented filter with weight w_0 on the central (first) sigma point and equal weights for the remaining sigma points, *RBP* is the Rao-Blackwellized particle filter with m particles.

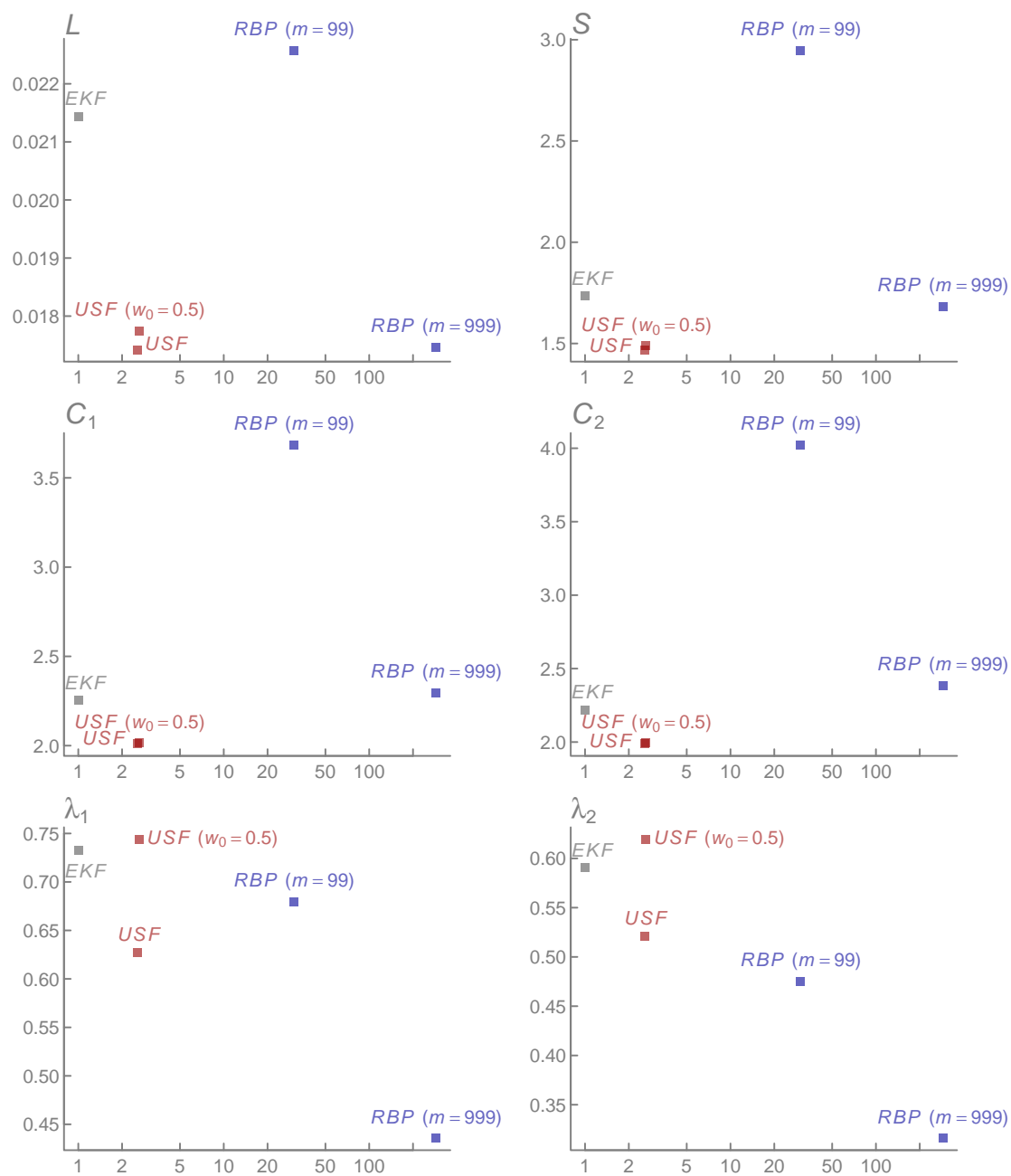


Figure 3. Accuracy versus computational cost tradeoff of alternative nonlinear filters. DNSS model with nine yields of maturities $\tau = 1/12, 6/12, 1, 2, 3, 5, 7, 10, 20$ years of 3000 observations. The horizontal axis is the log of computation time relative to the extended Kalman filter (EKF). The vertical axis is the average mean absolute scaled errors of the filtered states over 999 simulated draws of 3000 observations. L is the level, S the slope, C_1 , C_2 curvature factors with decay parameters λ_1 , λ_2 , respectively. USF is the unscented filter with equal weights on every sigma point, $USF (w_0 = 0.5)$ is the unscented filter with weight w_0 on the central (first) sigma point and equal weights for the remaining sigma points, RBP is the Rao–Blackwellized particle filter with m particles.

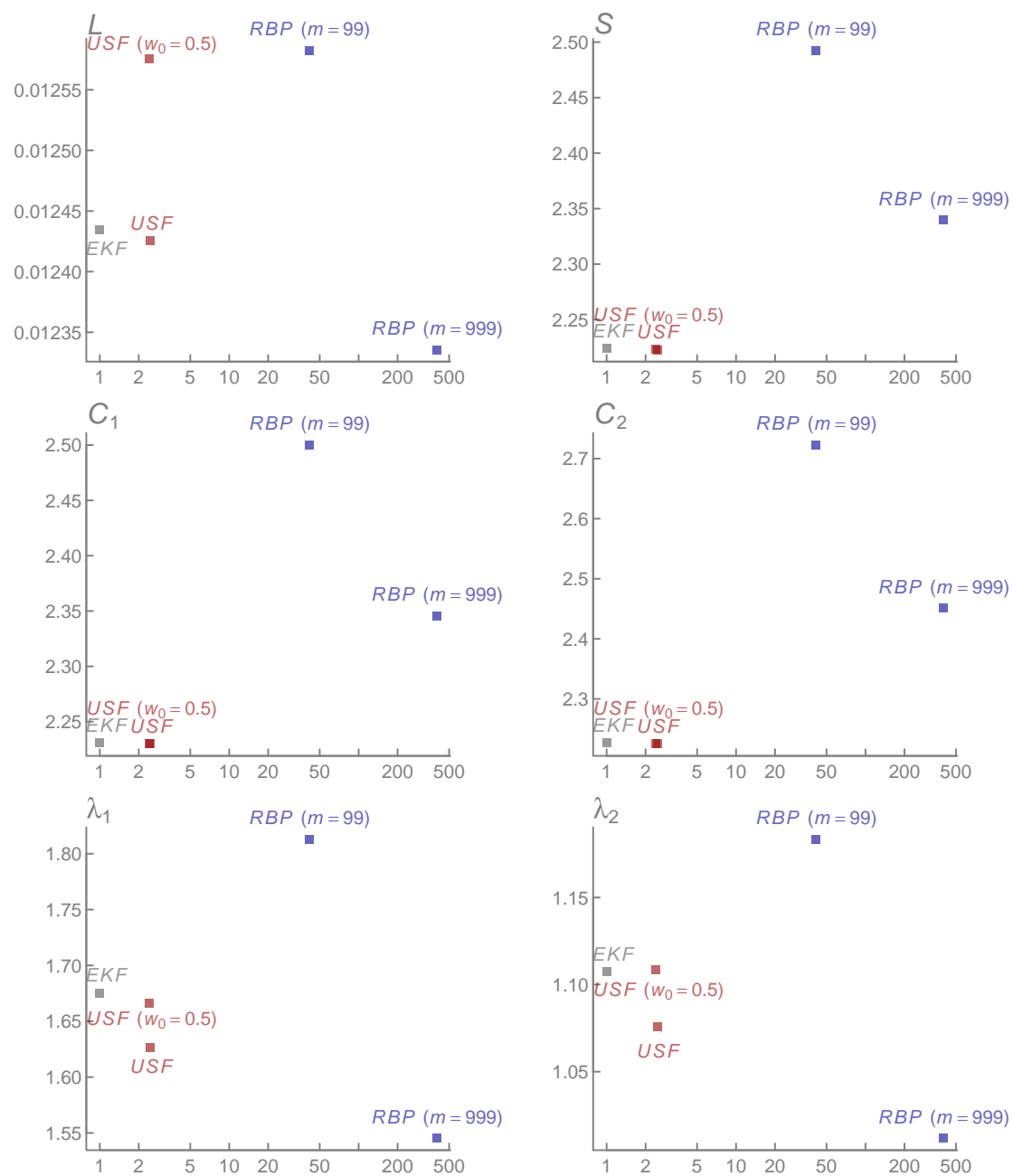


Figure 4. Accuracy versus computational cost tradeoff of alternative nonlinear filters. DNSS model with 13 yields of maturities $\tau = 1, 3/2, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25$ years of 3000 observations. The horizontal axis is the log of computation time relative to the extended Kalman filter (EKF). The vertical axis is the average mean absolute scaled errors of the filtered states over 999 simulated draws of 3000 observations. L is the level, S the slope, C_1 , C_2 curvature factors with decay parameters λ_1 , λ_2 , respectively. USF is the unscented filter with equal weights on every sigma point, $USF(w_0 = 0.5)$ is the unscented filter with weight w_0 on the central (first) sigma point and equal weights for the remaining sigma points, RBP is the Rao-Blackwellized particle filter with m particles.

5. Recovering Out-of-Sample Maturity Yields

This section analyzes real data in order to examine the predictability of out-of-sample maturity yields from dynamic term structure models with time-varying factors considered above.

5.1. Data

Two types of data are used for the empirical analysis. The first is daily yield data from the European Central Bank for the sample (6 September 2004 to 21 February 2019 (3701 observations) (Source: <http://sdw.ecb.europa.eu/browse.do?node=9691126>). These are known to be constructed from the fitted Nelson–Siegel–Svensson model. They provide two sets of estimated values of the six factors L , S , C_1 , C_2 , λ_1 , λ_2 from a cross-section of bonds, one consisting of AAA rated bonds and one consisting of all ratings. The observed yields are computed from these factors as fitted values from the Nelson–Siegel–Svensson model. That is, $y_t = Z(\alpha_t)$, where $Z(\alpha_t)$ is defined below (2) and it is equivalent to evaluating (1). Because the latent factors are observable in this data set, I can directly assess the in-sample accuracy of the filtered factors.

The second type of data come from central banks known to use spline based methods. The daily data are from Bank of Canada (4 January 2000 to 31 December 2018, 4719 observations), Bank of England (4 January 2000 to 31 December 2018, 4801 observations), and U.S. Department of Treasury (31 July 2001 to 31 December 2018, 4357 observations). All data are nominal yields and they are publicly available (Sources: Bank of Canada (<https://www.bankofcanada.ca/rates/interest-rates/bond-yield-curves/>), Bank of England (<https://www.bankofengland.co.uk/statistics/yield-curves>), U.S. Department of Treasury (<https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield>)). As mentioned in the introduction, data at the short- and long- end of the term structure are frequently missing in these data sets. This motivates examining the recoverability of out-of-sample maturity yields from the estimated dynamic term structure models. Although data are made publicly available, as far as I can tell, none of these central banks provide sufficient information to reproduce the yield curve data.

Figure 5 shows the time series of Nelson–Siegel–Svensson factor data from the European Central Bank. The level L and decay λ_1 , λ_2 parameters appear to be restricted to be positive, but not the slope S and curvature C_1 , C_2 factors. The two curvature and decay parameters are plotted together to highlight the weak identification problem. For the post-2015 period, the two decay factors nearly overlap indicating weak identification during this period. Another symptom of weak identification is the several ‘spikes’ or ‘jumps’ in the factor estimates. One cannot expect the DNSS model to accurately recover the latent factor series, particularly the curvature C_1 , C_2 and decay λ_1 , λ_2 factors, for this data set.

Table A1 in the online Appendix shows the results of fitting a univariate AR(1) model to each factor series from the European Central Bank, as assumed in the DNSS specification with diagonal T . The estimated AR(1) parameters are all above 0.98, with R^2 all above 0.96.

5.2. Estimation

I first need to estimate model parameters to predict out-of-sample maturity yields. For each data set, I divide the sample of maturities into two groups, one to be used for estimation and the other to be used for out-of-sample prediction. The grouping is somewhat arbitrary but the general rule for the in-sample maturities is to use the range of maturities commonly used in the existing literature with no missing values. For the out-of-sample maturities, I use maturities from both extreme ends of the term structure that typically have missing values. I also use one maturity not used in estimation that is within the maturity range used for estimation.

Figure 6 graphically summarizes the in-sample and out-of-sample maturities used in the empirical analysis. Because the European Central Bank (ECB) makes the underlying factor data available, we can choose any (reasonable) maturity and recover the yield data as fitted values, as described above. Two sets of in-sample maturities are used for the ECB data, one with nine maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ and one with thirteen maturities $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$. These are the maturities used in the simulation experiment in Section 4. These two choices are made in order to make them similar to the maturities

available from the other central banks (τ_9 for U.S. Department of Treasury data and τ_{13} for Bank of Canada and Bank of England data).

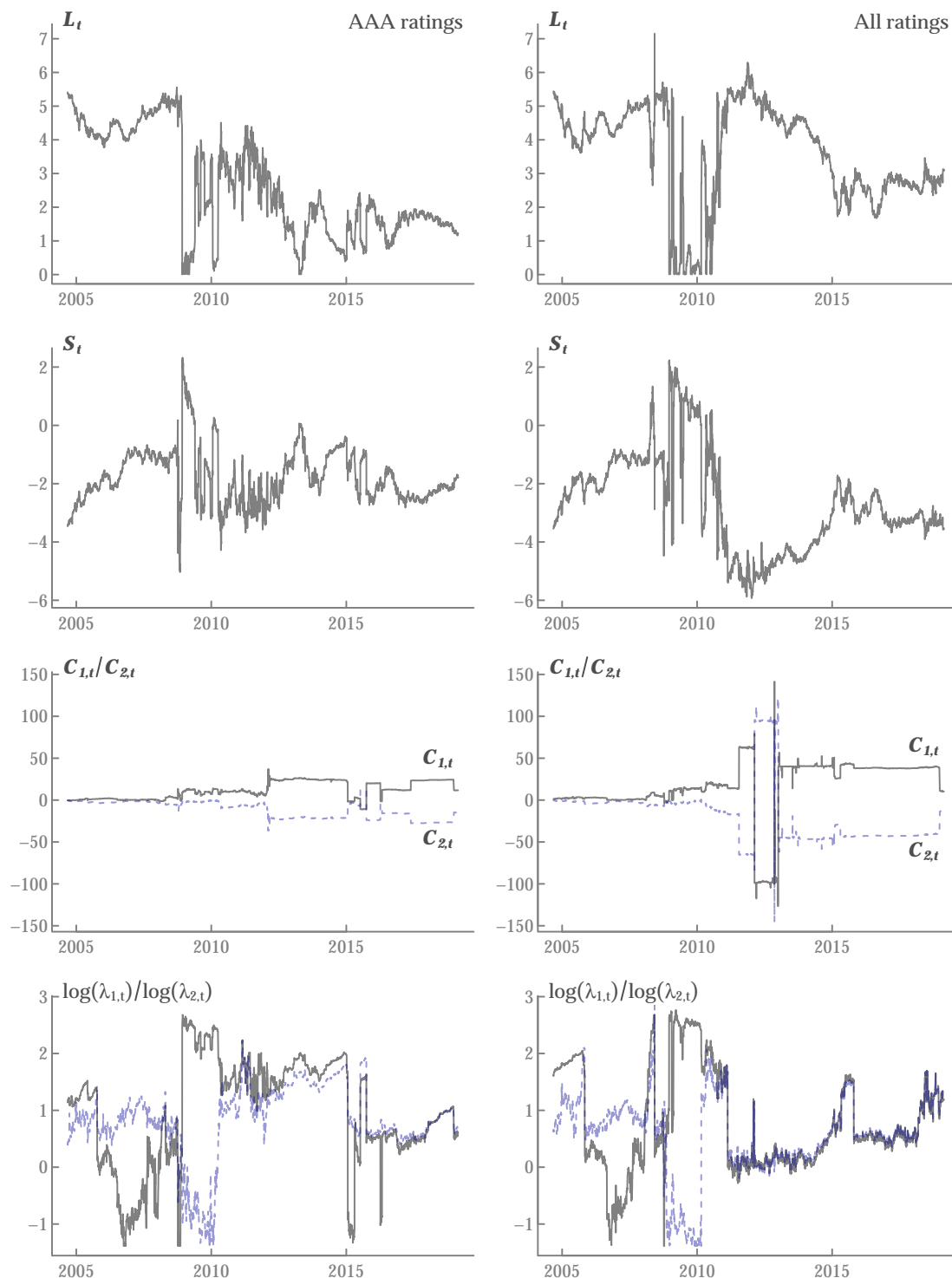


Figure 5. Nelson-Siegel-Svensson factors from the European Central Bank. The left side panels are for factors estimated from AAA bonds and the right side panels for factors estimated from all ratings. L_t is the level, S_t the slope, $C_{1,t}$, $C_{2,t}$ curvature factors with decay parameters $\lambda_{1,t}$, $\lambda_{2,t}$, respectively. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

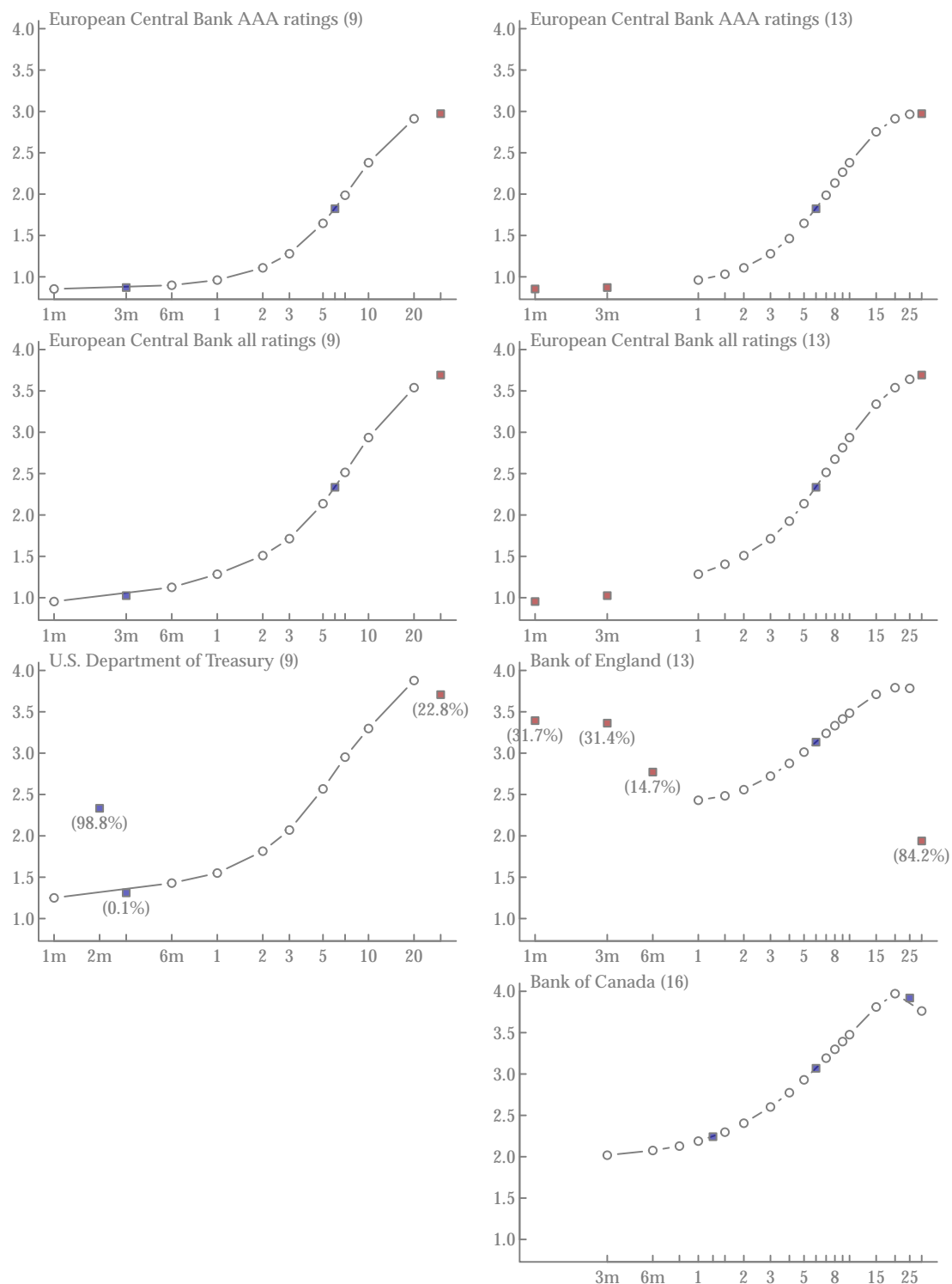


Figure 6. Yield curve maturities used for in-sample estimation and out-of-sample prediction. The vertical axis is the average yield (annualized percent) over the estimation and prediction samples for each maturity. The horizontal axis is maturity τ in log-scale for readability. The gray circles indicate in-sample maturities; the square filled boxes are out-of-sample maturities not used in estimation. Numbers in parentheses are missing observations (as percentage of the sample) for out-of-sample maturities, if there are any.

The parameters of the model are estimated by numerically maximizing the gaussian log-likelihood function (4) while using the extended Kalman filter. The likelihood is numerically maximized using

a gradient based optimizer. Because of the large number of parameters to estimate, especially for the *DNSS* model, finite difference methods to numerically approximate the gradients are not only computationally costly, but can also be inaccurate. The scores are evaluated using algorithmic differentiation (The R package TMB (<https://cran.r-project.org/package=TMB>) was used to evaluate the likelihood and scores). Because of the weak identification problem discussed in Section 3.3, numerical optimization is sensitive to the choice of starting parameter values. Not only do they often fail to converge, but when they do converge they do not always converge to similar parameter estimates. The numerical optimizer is run through a number of randomly selected starting values and the estimates are taken among results that converged with the highest likelihood value.

The parameter estimates are reported in Tables A2–A8 of the online Appendix. The QML standard errors are computed by evaluating both the scores and Hessian with algorithmic differentiation. As another symptom of weak identification, the Hessian evaluated at the parameter estimates are sometimes singular. For such cases, the reported standard errors use the pseudo-inverse of the Hessian. Consistent with the least squares estimates fitted to the underlying factor data, the estimated latent series are all highly persistent with AR(1) parameter close to the nonstationary boundary of one (the AR(1) parameters were constrained to be stationarity when maximizing the likelihood).

For the European Central Bank data, we can compare how well the filtered state series using these estimates can recover the factor data. Figures 7 and 8 compare the ‘true’ α_t and filtered $a_{t|t}$ state series for the AAA rating data (Figures A1 and A2 in the Appendix A show the comparisons for the data based on all ratings). Consistent with the simulation results, the first two factors L and S are reasonably well recovered, especially when using thirteen maturities τ_{13} . The two curvature factors C_1 , C_2 are tracked reasonably well for the first half of the sample, but not for the second half after some spikes or jumps in the true data. The two decay parameters are close to each other in the second half of the sample and C_1 , C_2 are likely to be highly collinear and weakly identified during this period, as observed in Figure 5. The two decay factors are never well tracked in all cases, highlighting the difficulty of identifying these factors from the observed yields.

5.3. Prediction

Figures 7 and 8 show the difficulty of recovering the underlying state series of the *DNSS* model due to weak identification. However, this does not necessarily imply that predictions from these model may not perform well. Under weak identification, there may be several alternative parameter configurations that, nonetheless, produce similar predictions.

The focus is on the cross-sectional predictability rather than the more commonly tested time series predictability in order to examine the predictive performance of these models. This focus on the cross-sectional predictability is motivated by the fact that data for very short- and long-ends of the term structure are often missing, as can be seen in Figure 6.

The cross-sectional predictions for yields of out-of-sample maturity τ are obtained from the measurement Equation (2) using the filtered state series $a_{t|t}$ as

$$\hat{y}_t^{(\tau)} = Z^{(\tau)}(a_{t|t}), \quad a_{t|t} = E_t[\alpha_t]$$

\hat{y} uses the filtered series rather than the one-step predictions $a_{t|t-1} = E_{t-1}[\alpha_t]$ or the smoothed series $a_{t|n} = E_n[\alpha_t]$. The smoothed series uses information from future observations and it is subject to ‘look ahead’ bias. The one-step prediction may be a better choice to mimic real-time, or online, prediction situations than the filtered series. The filtered series are used, so that we can compare the model based predictor with a ‘naive’ interpolated predictor. If there are maturities τ_1 , τ_2 used for estimation that bracket the out-of-sample maturity τ_0 , such that $\tau_1 < \tau_0 < \tau_2$, a simple natural predictor is the linearly interpolated value

$$\hat{y}_t^{(\tau_0)} = \frac{\tau_2 - \tau_0}{\tau_2 - \tau_1} y_t^{(\tau_1)} + \frac{\tau_0 - \tau_1}{\tau_2 - \tau_1} y_t^{(\tau_2)}$$

Because these interpolated values use information up to t , they are comparable to forecasts that are based on filtered series rather than one-step predictions. The main advantage of the model based predictor over interpolation is that they apply even when the out-of-sample maturity have no bracketing in-sample maturities. Extrapolating outside the in-sample maturity range is more problematic than interpolating within the in-sample maturity range.

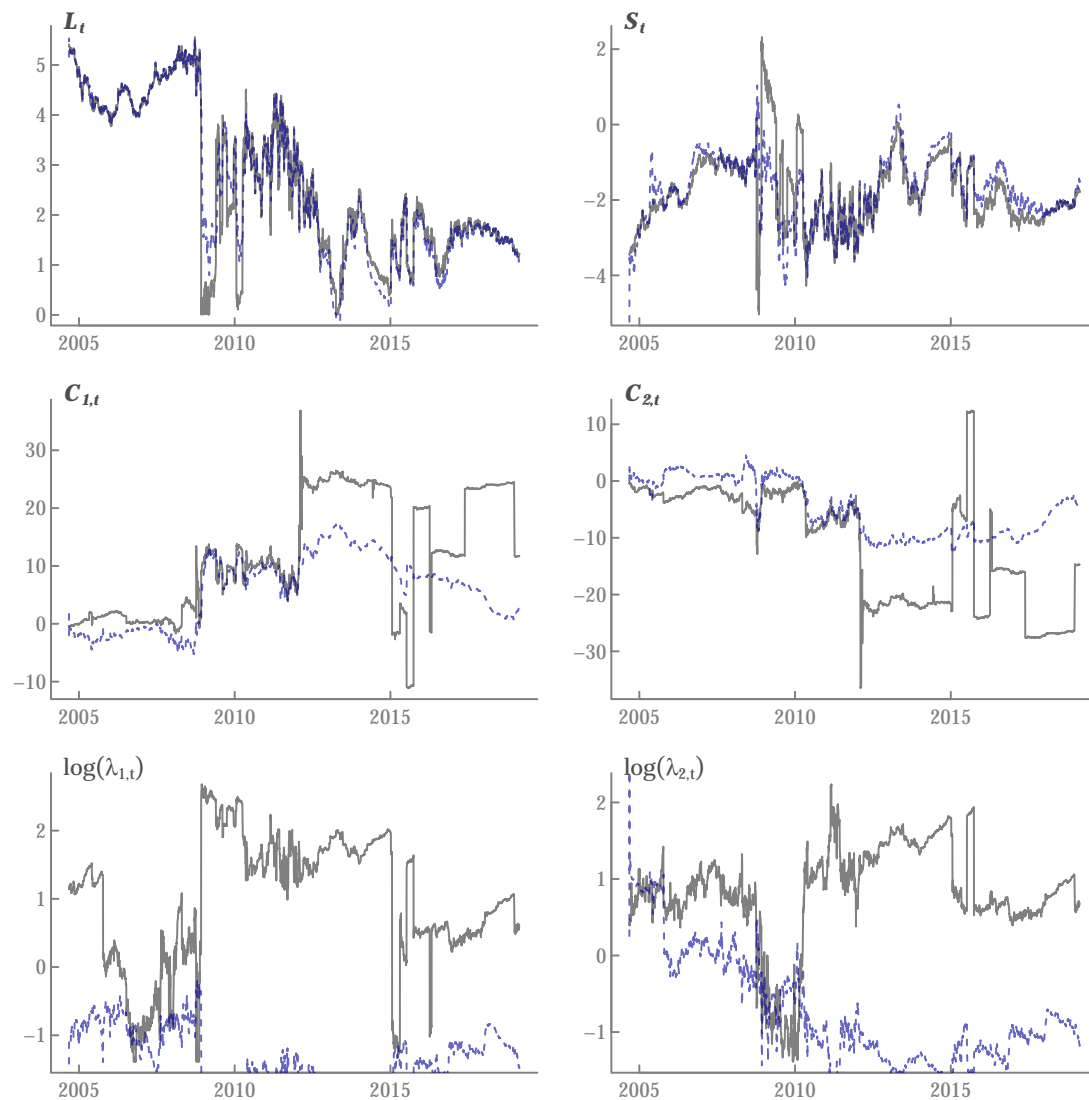


Figure 7. Nelson-Siegel-Svensson factors (AAA ratings) from the European Central Bank. L_t is the level, S_t the slope, $C_{1,t}$, $C_{2,t}$ curvature factors with decay parameters $\lambda_{1,t}$, $\lambda_{2,t}$, respectively. The gray solid lines are the actual factors and the blue dashed lines are filtered estimates from the extended Kalman filter. The parameters of the state space model were estimated using 13 zero yield curves with maturities $\tau = 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25$ (years). The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

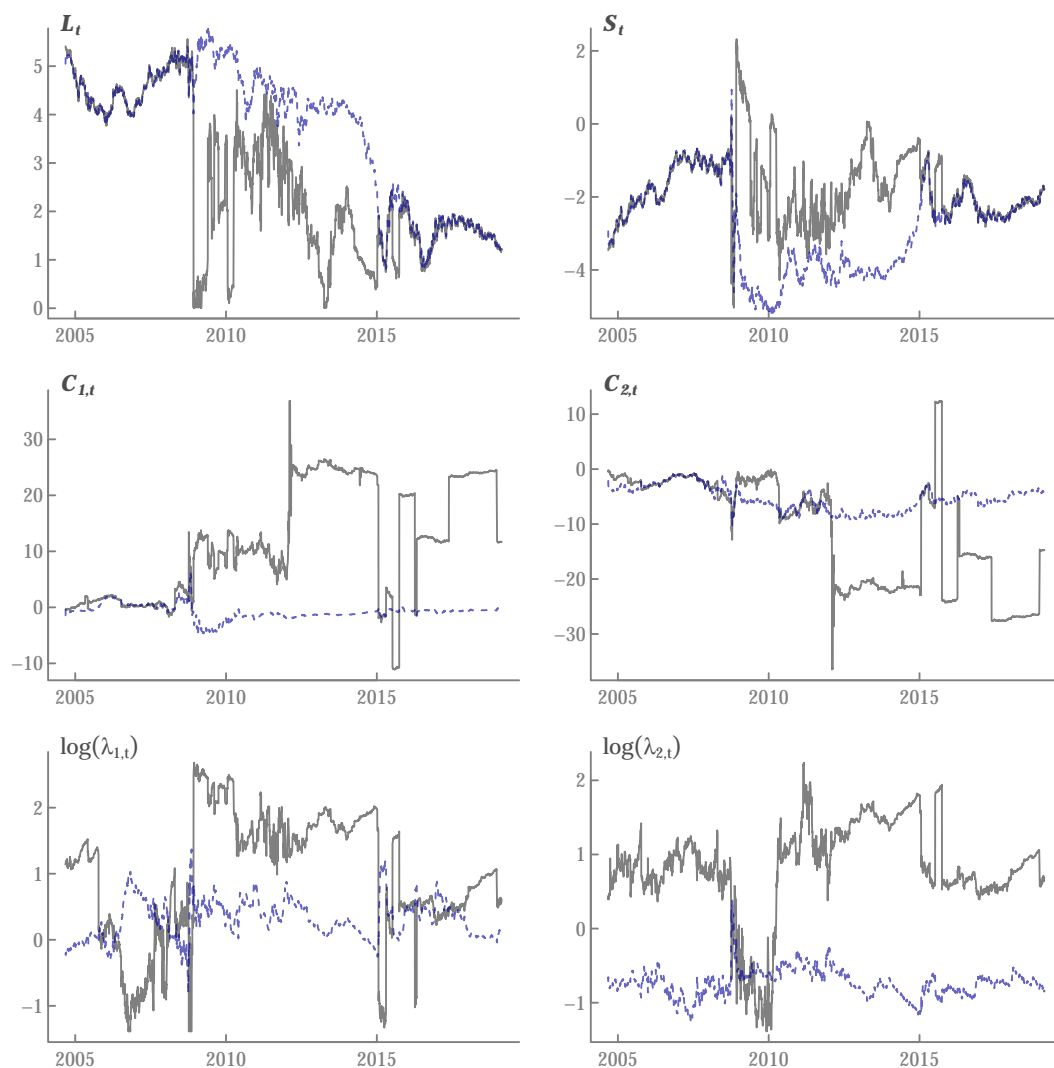


Figure 8. Nelson–Siegel–Svensson factors (AAA ratings) from the European Central Bank. L_t is the level, S_t the slope, $C_{1,t}$, $C_{2,t}$ curvature factors with decay parameters $\lambda_{1,t}$, $\lambda_{2,t}$, respectively. The gray solid lines are the actual factors and the blue dashed lines are filtered estimates from the extended Kalman filter. The parameters of the state space model were estimated using 9 zero yield curves with maturities $\tau = 1/12, 0.5, 1, 2, 3, 5, 7, 10, 20$ (years). The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

The model based predictors $\hat{y}_t^{(\tau)}$ that are based on the filtered state series are subject to sampling errors. The conditional mean squared prediction error can be computed as

$$\begin{aligned} E_t[(\hat{y}_t^{(\tau)} - y_t^{(\tau)})(\hat{y}_t^{(\tau)} - y_t^{(\tau)})^\top] &= E_t[(Z^{(\tau)}(a_{t|t}) - Z^{(\tau)}(\alpha_t))(Z^{(\tau)}(a_{t|t}) - Z^{(\tau)}(\alpha_t))^\top] + E_t[\epsilon_t \epsilon_t^\top] \\ &\approx \frac{\partial Z^{(\tau)}}{\partial \alpha^\top} \text{Cov}_t(a_{t|t}) \frac{\partial Z^{(\tau)}}{\partial \alpha} + H_t \\ &= \frac{\partial Z^{(\tau)}}{\partial \alpha^\top} A_{t|t} \frac{\partial Z^{(\tau)}}{\partial \alpha} + H_t \end{aligned}$$

The second line approximation is due to the delta method. There are two difficulties with assessing the size of sampling error in the predictions. First, the above expression does not account for parameter uncertainty, the fact that the filtered series are computed using estimated parameter values that are themselves subject to sampling error. Second, there is no natural estimator for the innovation variance H for out-of-sample maturities, since H is associated with each in-sample maturity.

An alternative approach to deal with the latter problem is to turn the prediction problem into an estimation problem. This can be accommodated within the state-space model by including the partially observed yields in the estimation sample. The state-space filter can then fill in the missing values (Section 4.10 in [28]). This approach is left for future work, including how the fraction of missing observations in the sample affects the performance as compared to the prediction approach used in this paper.

Because of these difficulties, instead of using formal statistical tests I use informal visual diagnostics to check the cross-sectional predictability of the model based forecasts. For time series forecast evaluation, it is common practice to specify a loss function and compare certain moments of the forecast error distribution, such as the mean absolute error (for absolute error loss) or root mean squared error (for quadratic loss). A recent work by Diebold and Shin [29], Jin et al. [30] proposed an approach that does not depend on the choice of loss function. Rather than focus on certain moments of the error distribution, this approach checks how ‘close’ the forecast error cumulative distribution function (cdf) is to that from a perfect forecast. The cdf of a perfect forecast jumps vertically from 0 to 1 at $e = 0$ where e is the support of the forecast errors. Jin et al. [30] construct a formal test that measures the closeness of the two cdfs. Because of the difficulties with formal testing alluded to above, in this application I plot the empirical cdfs of the forecast errors and visually compare how close they are to the cdf of perfect forecasts.

Figures 9–11 compare the empirical cdfs of the prediction errors (The European Central Bank data for all ratings is given in Figure A3 of the online Appendix). For the European Central Bank data in Figure 9, it is visually apparent that the predictions from the *DNSS* model is closest to the vertical line at $e = 0$ for most maturities out-of-sample. Figure A4 in the online Appendix compares the kernel density estimates of the error distributions. For most out-of-sample maturities the density for the *DNSS* error distribution has least bias and variance. Despite the weak identification problem with *DNSS*, this result is not surprising, since we know that the data are generated from the *DNSS* model for the European Central Bank data.

Of more interest is for the data estimated from spline based methods that are shown in Figures 10 and 11. (Figures A6 and A8 in the online Appendix compare the corresponding kernel density estimates of the distribution of prediction errors.) For these data, the results depend on the data and maturity. For the Bank of Canada and U.S. Department of Treasury data shown in Figure 10, the interpolation method is often the best with *DNSS* a close second. However, the interpolation method can only be used for non-extreme maturities within the range of maturities used for estimation. For the Bank of England data presented in Figure 11 there is not much difference (at least visually) among the different models, except for *NSS*, at the short-end of the term structure. No model performs well at the long-end, although one must keep in mind that these comparisons are based on limited amount of data due to a large fraction of missing observations; see Figure 6.

Figure 12 shows the difference in prediction mean absolute errors (MAE) between *DNSS* and the other models to illustrate how loss function based summary predictive performance measures compare with the distribution based measures used above. A negative difference indicates better prediction from the *DNSS* model than from the comparison model. These results correspond to the cross-sectional out-of-sample forecasts for the spline based data in Figures 10 and 11. The confidence intervals for the difference in MAE were constructed from the Diebold and Mariano [31] statistic with heteroskedasticity and autocorrelation robust (HAR) standard errors. For maturities inside the range of maturities used for estimation, a simple linear interpolation generally performs better. However, maturities at both ends of the term structure that require extrapolation are the ones that are often missing. For these cases the *DNSS* model generally performs better than the more restrictive parametric models.

Figure A9 in the online Appendix shows the difference in prediction mean squared errors (MSE). These results generally agree with those based on MAE in Figure 12. For the Bank of England data

for maturities $\tau = 1/2$ and $\tau = 6$, some of the differences are not statistically significant using MSE. This illustrates the sensitivity of performance ranking to the choice of loss function.

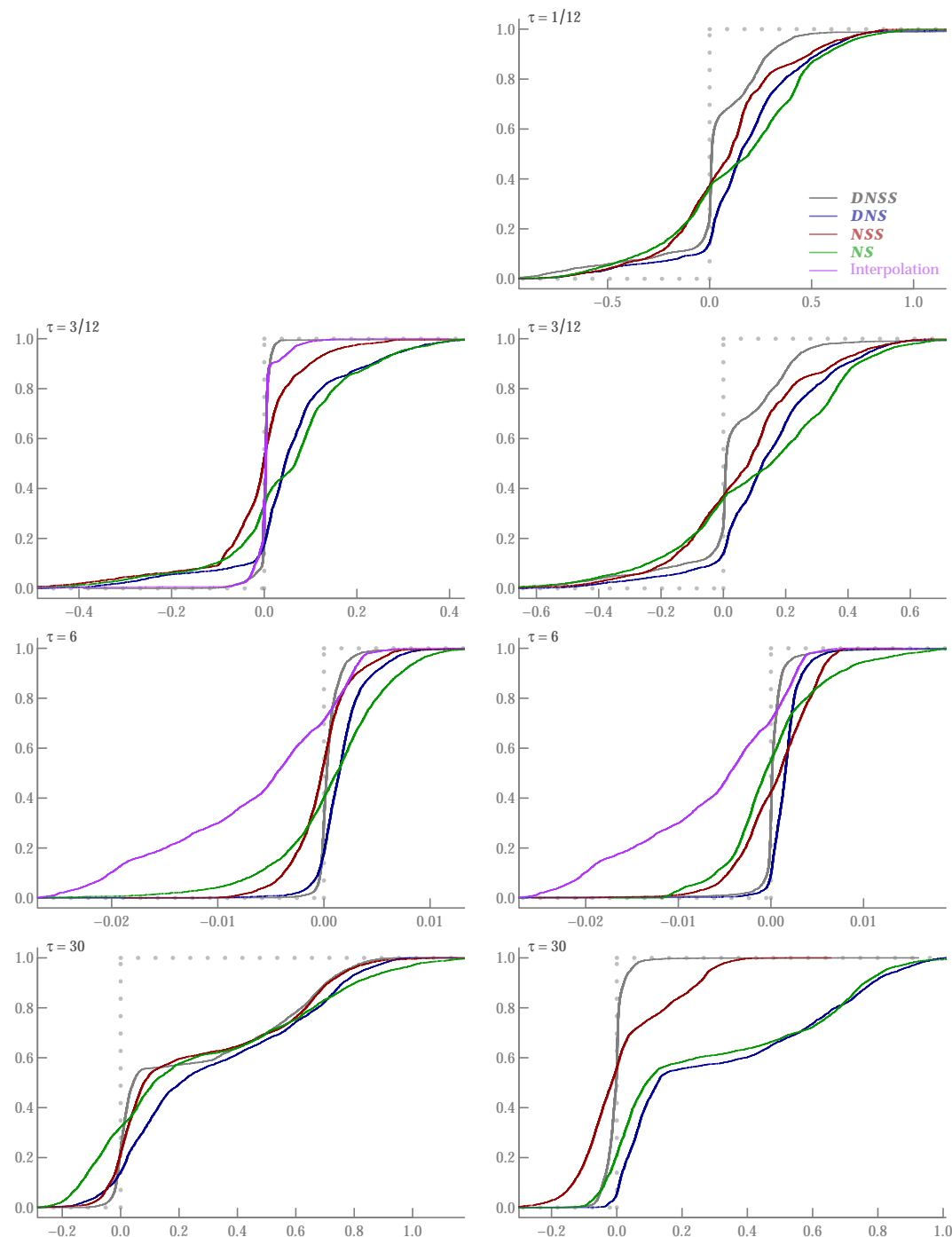


Figure 9. Empirical cdf of prediction errors. Predictions from European Central Bank (AAA ratings) estimates with maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ (left panel) and $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$ (right panel). DNSS is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, DNS is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, NSS is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, NS is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The distribution of a perfect prediction with no error would be a vertical line at $e_t = 0$ indicated as a dotted line. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

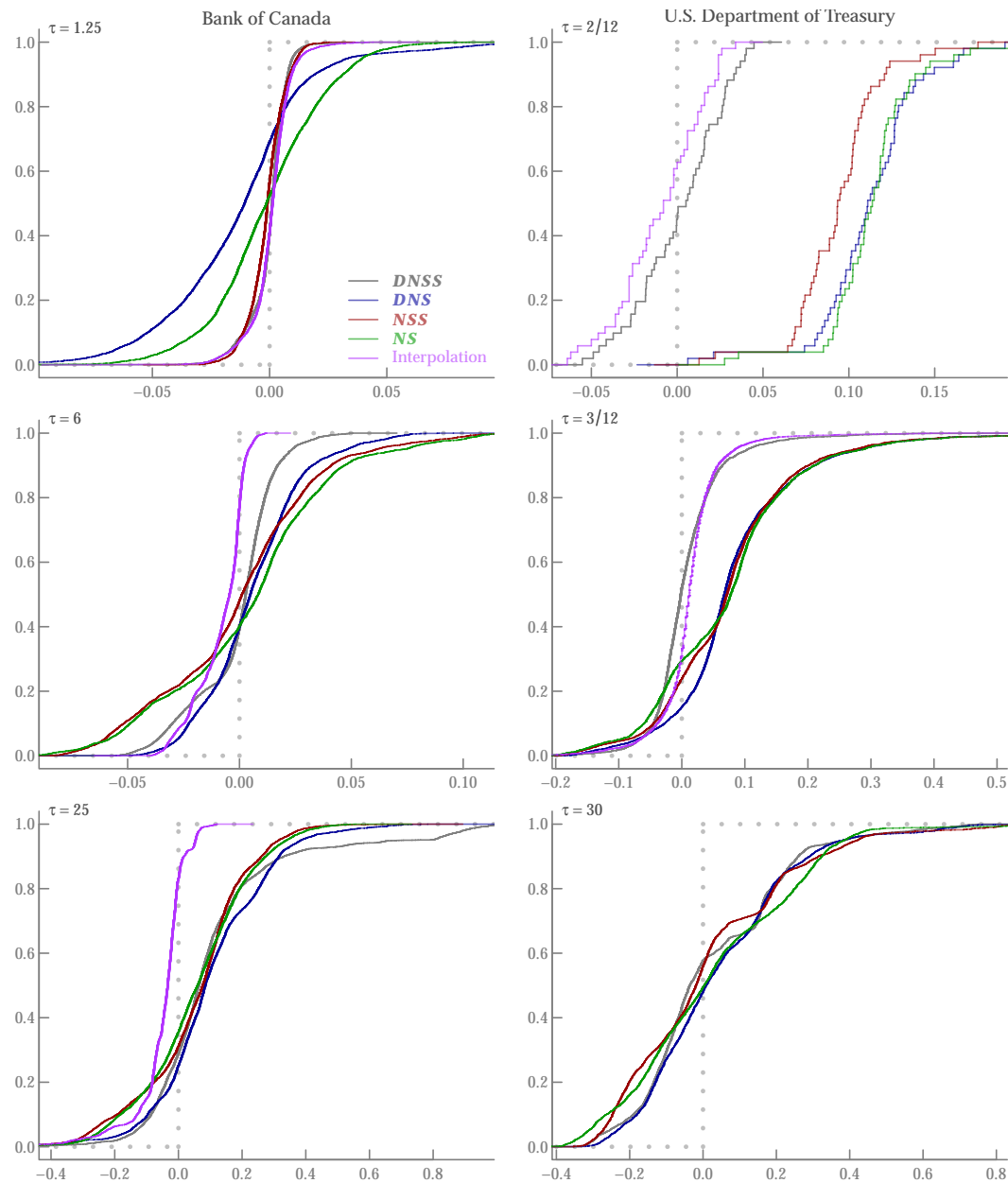


Figure 10. Empirical cdf of prediction errors. Predictions from Bank of Canada data with maturities $\tau_{16} = (3/12, 6/12, 0.75, 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30)$ (left panel) and U.S. Department of Treasury with maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ (right panel). *DNSS* is the dynamic Nelson–Siegel–Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson–Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson–Siegel–Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson–Siegel specification with a single decay parameter λ_1 assumed to be fixed. The distribution of a perfect prediction with no error would be a vertical line at $e_t = 0$ indicated as a dotted line. The daily sample is 4 January 2000 to 31 December 2018 (4719 observations) for Bank of Canada (left panel) and 31 July 2001 to 31 December 2018 (4357 observations some of which may be missing) for U.S. Department of Treasury (right panel).

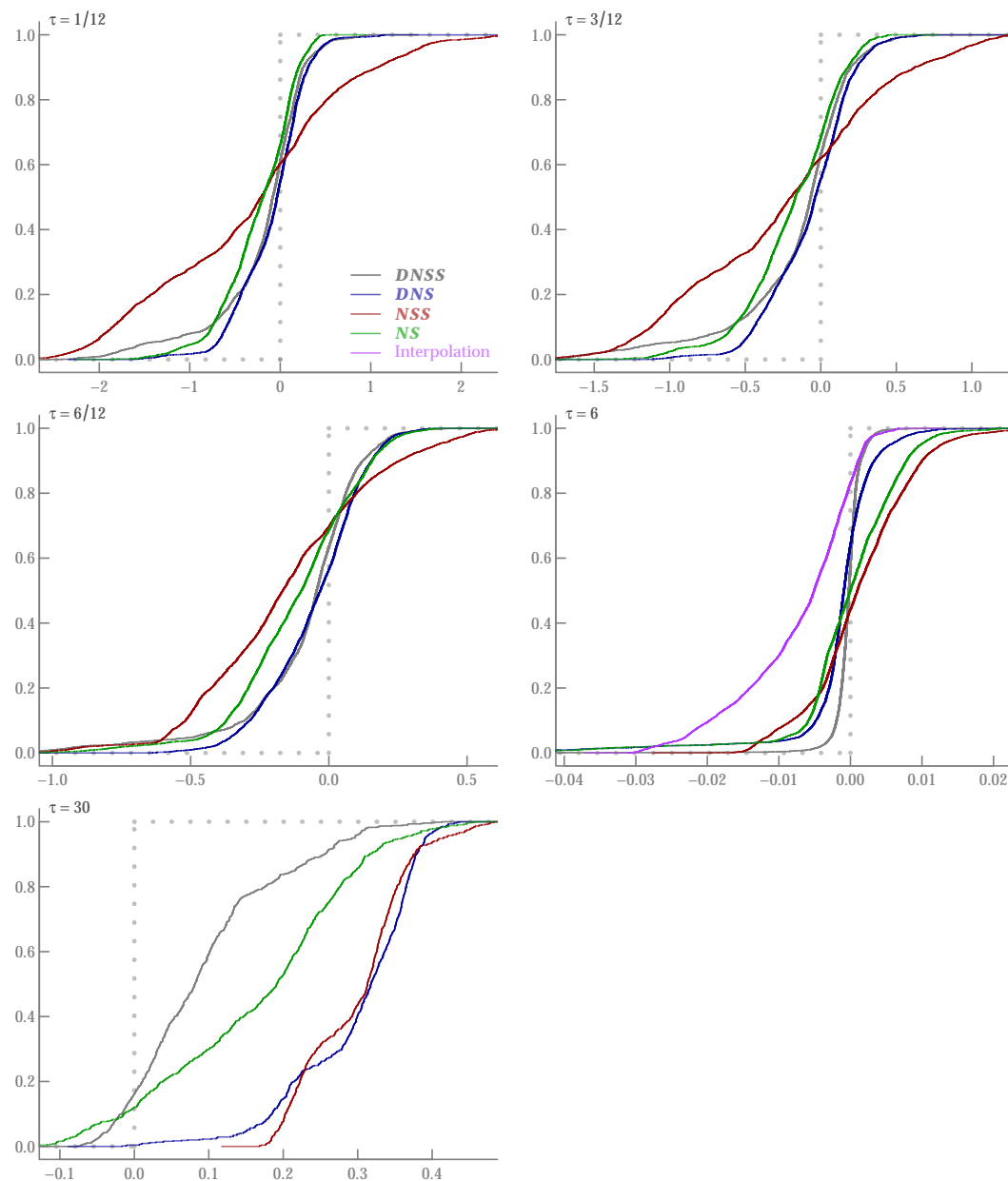


Figure 11. Empirical cdf of prediction errors. Predictions from Bank of England data with maturities $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$. *DNSS* is the dynamic Nelson–Siegel–Svensson specification, where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson–Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson–Siegel–Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson–Siegel specification with a single decay parameter λ_1 assumed to be fixed. The distribution of a perfect prediction with no error would be a vertical line at $e_t = 0$ indicated as a dotted line. The daily sample is 1 April 2000 to 31 December 2018 (4801 observations some of which may be missing).

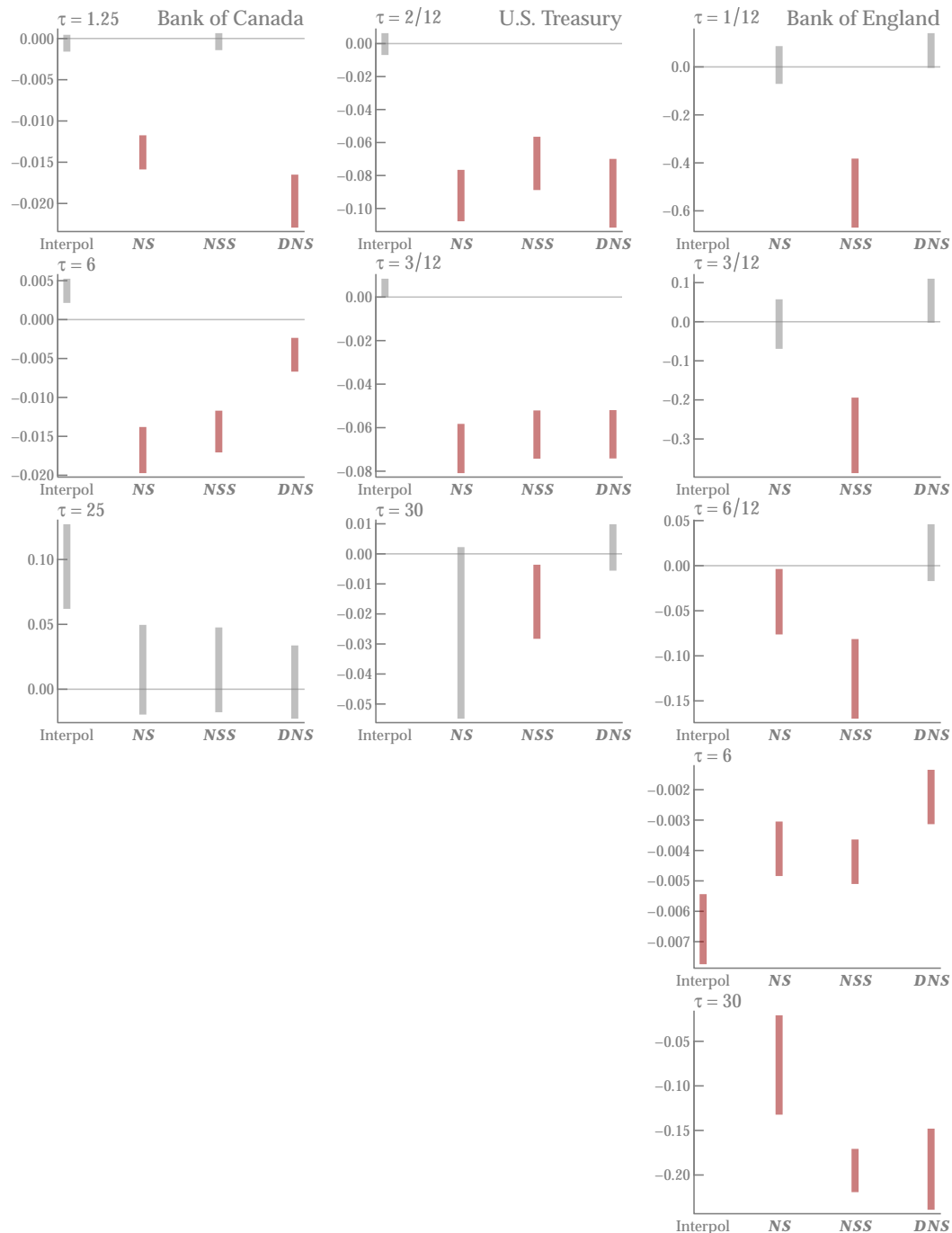


Figure 12. Prediction mean absolute error (MAE) differences. Plotted are 0.95 confidence intervals for the difference in MAE between *DNSS* and other models. A negative difference indicates better forecasts for *DNSS* relative to the comparison model. Red intervals indicate statistical significance at size 0.05. *DNSS* is the dynamic Nelson–Siegel–Svensson specification, where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson–Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson–Siegel–Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson–Siegel specification with a single decay parameter λ_1 assumed to be fixed.

6. Concluding Remarks

This research was motivated by the need to recover missing yield data at the extreme ends, very short- and long- maturities, of the term structure. I sincerely hope that more central banks follow the

lead taken by the European Central Bank and provide sufficient information in order to reproduce the yield estimates. The sooner the need that motivated this research becomes obsolete, the better.

However, currently, data produced by spline based methods employed by several major central banks remain in a black box. For researchers who do not have access to the underlying bond data, the class of parametric dynamic term structure models considered in this paper could be used to infer data that are often missing at the extreme ends of the term structure. The performance of alternative parametric specifications is somewhat mixed and data dependent. The advice is to carefully consider how predictions from alternative specifications differ from each other and consider the tradeoff between parsimony or accuracy and in-sample overfitting. The result that more complex specifications do not necessarily produce better out-of-sample predictions is consistent, for example, with the assessment in Nyman-Andersen [1] based on the cross-section of bond prices.

There are several directions in which the parametric dynamic term structure approach can be further developed. Due to computational constraints, the specifications considered in this paper have restricted the state variable dynamics to the diagonal AR(1) specification. A general VAR(1) specification would not only result in estimating a much larger number of parameters, but also introduce additional issues, such as constraining the estimates, to remain in the stationarity region [6]. If computing power were not a constraint (such as with access to cloud computing clusters), estimation using alternative nonlinear filters can also be pursued.

Another potentially important issue not addressed in the current paper is the effect of structural break or jumps in the underlying state series. The time series plot of the factors presented in Figure 5 show several spikes or jumps, especially around the financial crisis period in 2008. The state-space model could be generalized in order to accommodate such potential breaks following the approach in Andrieu et al. [32], Nemeth et al. [33].

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Appendix A.1. Linearization for the Extended Kalman Filter

For the DNSS specification

$$Z(\alpha) = 1_m \alpha_1 + f(e^{\alpha_5}) \alpha_2 + g(e^{\alpha_5}) \alpha_3 + g(e^{\alpha_6}) \alpha_4$$

$$f(\lambda) = \frac{1}{\lambda \tau} (1 - e^{-\lambda \tau}), \quad g(\lambda) = f(\lambda) - e^{-\lambda \tau}$$

where $\alpha_t = (L_t, S_t, C_{1t}, C_{2t}, \log \lambda_{1t}, \log \lambda_{2t})$. Take Taylor expansion around $\alpha_t = a_t$ to get

$$Z(\alpha) \approx Z(a) + \frac{\partial Z(a)}{\partial \alpha^\top} (\alpha - a)$$

where $\partial Z(\alpha) / \partial \alpha^\top$ is the $m \times 6$ Jacobian matrix

$$\frac{\partial Z(\alpha)}{\partial \alpha^\top} = (1_m, f(e^{\alpha_5}), g(e^{\alpha_5}), g(e^{\alpha_6}), (\frac{\partial f(\lambda_1)}{\partial \lambda} \alpha_2 + \frac{\partial g(\lambda_1)}{\partial \lambda} \alpha_3) e^{\alpha_5}, \frac{\partial g(\lambda_2)}{\partial \lambda} \alpha_4 e^{\alpha_6})$$

$$\frac{\partial f(\lambda)}{\partial \lambda} = \frac{1}{\lambda^2 \tau} ((1 + \lambda \tau) e^{-\lambda \tau} - 1) = -\frac{g(\lambda)}{\lambda}$$

$$\frac{\partial g(\lambda)}{\partial \lambda} = \frac{\partial f(\lambda)}{\partial \lambda} + \tau e^{-\lambda \tau}$$

with $\lambda_1 = e^{\alpha_5}$, $\lambda_2 = e^{\alpha_6}$.

For the *DNS* specification

$$Z(\alpha) = 1_m \alpha_1 + f(e^{\alpha_4}) \alpha_2 + g(e^{\alpha_4}) \alpha_3$$

where $\alpha_t = (L_t, S_t, C_t, \log \lambda_t)$. The $m \times 4$ Jacobian matrix $\partial Z(\alpha) / \partial \alpha^\top$ is

$$\frac{\partial Z(\alpha)}{\partial \alpha^\top} = (1_m, f(e^{\alpha_4}), g(e^{\alpha_4}), (\frac{\partial f(\lambda)}{\partial \lambda} \alpha_2 + \frac{\partial g(\lambda)}{\partial \lambda} \alpha_3) e^{\alpha_4})$$

with $\lambda = e^{\alpha_4}$.

Appendix A.2. Unscented Kalman Filter

s -dimensional target distribution with first two moments $x \sim (\bar{x}, P_{xx})$. The $2s + 1$ sigma points x_i and weights w_i that match the first two moments of x are [23,34,35]

$$x_0 = \bar{x}, \quad x_i = \bar{x} + \lambda L_{i,i}, \quad x_{i+s} = \bar{x} - \lambda L_{i,i} \\ w_i = w_{i+s} = \frac{1 - w_0}{2s}$$

for $i = 1, \dots, s$ where $L_{i,i}$ is the i -th column of L such that $LL^\top = P_{xx}$. To determine λ , w_0 , set $w_0 = k/(s+k)$ for some k , a tuning parameter. Then the second moment is matched with $\lambda = \sqrt{s+k}$. If target x is gaussian, set $s+k = 3$ or $k = 3-s$. This may result in negative weights for $s > 3$. Equivalently, set w_0 as tuning parameter with $k = w_0/(1-w_0)s$ and $\lambda = \sqrt{s/(1-w_0)}$.

Updating step of filter: estimate $a_{t|t}$, $P_{t|t}$ given y_t , $a_t = a_{t|t-1}$, $P_t = P_{t|t-1}$ with sigma points and weights

$$x_{0,t} = a_t, \quad x_{i,t} = a_t + \lambda L_{i,i}, \quad x_{i+s,t} = a_t - \lambda L_{i,i} \\ w_0 = \frac{k}{s+k}, \quad w_i = w_{i+s} = \frac{1 - w_0}{2s}$$

More specifically, compute

$$y_{i,t} = Z_t(x_{i,t}), \quad \bar{y}_t = \sum_{i=0}^{2s} w_i y_{i,t}, \\ P_{av,t} = \sum_{i=0}^{2s} w_i (x_{i,t} - a_t)(y_{i,t} - \bar{y}_t)^\top, \quad P_{vv,t} = \sum_{i=0}^{2s} w_i ((y_{i,t} - \bar{y}_t)(y_{i,t} - \bar{y}_t)^\top + H_t(x_{i,t}))$$

and update

$$a_{t|t} = a_t + P_{av,t} P_{vv,t}^{-1} (y_t - \bar{y}_t) \\ P_{t|t} = P_t - P_{av,t} P_{vv,t}^{-1} P_{av,t}^\top$$

Prediction step of filter: estimate a_{t+1} , P_{t+1} given $a_{t|t}$, $P_{t|t}$ with sigma points (same weights)

$$x_{0,t} = a_{t|t}, \quad x_{i,t} = a_{t|t} + \lambda L_{i,i}, \quad x_{i+s,t} = a_{t|t} - \lambda L_{i,i}$$

and predict

$$a_{t+1} = \sum_{i=0}^{2s} w_i T_t(x_{i,t}) \\ P_{t+1} = \sum_{i=0}^{2s} w_i ((T_t(x_{i,t}) - a_{t+1})(T_t(x_{i,t}) - a_{t+1})^\top + R_t(x_{i,t}) Q_t(x_{i,t}) R_t(x_{i,t})^\top)$$

For linear state equations, just apply the usual prediction step

$$a_{t+1} = c_t + T_t a_{t|t}, \quad P_{t+1} = T_t P_{t|t} T_t^\top + R_t Q_t R_t^\top$$

without approximation error (Rao-Blackwellization).

Unscented filter accurate to second order for the mean. Extended Kalman filter only accurate to first order for the mean. The variance estimate accurate to second order for both filters. If $k < 0$, the prediction step may yield a non-positive definite covariance matrix.

Appendix A.3. Rao-Blackwellized Particle Filter

The material in this section largely taken from Schon et al. [26] for the reader's convenience. Write the state space model as

$$\begin{aligned}\lambda_{t+1} &= d_2 + T_{22}\lambda_t + T_{21}x_t + v_{2,t+1} \\ x_{t+1} &= d_1 + T_{12}\lambda_t + T_{11}x_t + v_{1,t+1} \\ y_t &= C(\lambda_t)x_t + e_t, \quad e \sim N(0, R)\end{aligned}$$

This is model 3 of (18), p. 2282 in Schon et al. [26] with $x_t^n = \lambda_t$, $x_t^\ell = x_t$, $f_t^n = d_2 + T_{22}\lambda_t$, $A_t^n = T_{21}$, $f_t^\ell = d_1 + T_{12}\lambda_t$, $A_t^\ell = T_{11}$, $h_t = 0$.

The conditional distributions for drawing the nonlinear state variables λ_t are (25), p. 2283 in [26]

$$\begin{aligned}p(y_t | \Lambda_t, Y_{t-1}) &\sim N(C_t \hat{x}_{t|t-1}, C_t P_{t|t-1} C_t^\top + R) \\ p(\lambda_{t+1} | \Lambda_t, Y_t) &\sim N(f_t^n + T_{21} \hat{x}_{t|t}, T_{21} P_{t|t} T_{21}^\top + Q_{22})\end{aligned}$$

and (12), p. 2281 in [26]

$$p(\Lambda_t | Y_t) = \frac{p(y_t | \Lambda_t, Y_{t-1}) p(\lambda_t | \Lambda_{t-1}, Y_{t-1})}{p(y_t | Y_{t-1})} p(\Lambda_{t-1} | Y_{t-1})$$

Filtering algorithm p. 2280 in [26].

- (1) Initialize. Draw $\lambda_{0|-1}^i \sim p(\lambda_0)$ for $i = 1, \dots, N$ particles. Set $x_{0|-1}^i = E[x_0]$, $P_{0|-1}^i = \text{Cov}(x_0)$.
- (2) Evaluate importance weights $q_t^i = p(y_t | \Lambda_t, Y_{t-1})$. Simulated contribution to log-likelihood is $\ell_t = \log(\sum_i q_t^i / N)$. Normalize $\tilde{q}_t^i = q_t^i / \sum_j q_t^j$.
- (3) Particle filter measurement update. Resample N particles with replacement $\Pr(\lambda_{t|t}^i = \lambda_{t|t-1}^j) = \tilde{q}_t^j$.
- (4a) Kalman filter measurement update (22), p. 2283 in [26]

$$\begin{aligned}M_t &= C_t P_{t|t-1} C_t^\top + R, & K_t &= P_{t|t-1} C_t^\top M_t^{-1} \\ \hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t (y_t - C_t \hat{x}_{t|t-1}), & P_{t|t} &= P_{t|t-1} - K_t M_t K_t^\top\end{aligned}$$

- (4b) Particle filter time update (prediction) $\lambda_{t+1|t}^i \sim p(\lambda_{t+1|t} | \Lambda_t^i, Y_t)$.
- (4c) Kalman filter time update (23), p. 2283 in [26]

$$\begin{aligned}\bar{A}^\ell &= T_{11} - Q_{21} Q_{22}^{-1} T_{21}, & \bar{Q}^\ell &= Q_{11} - Q_{21} Q_{22}^{-1} Q_{21}^\top \\ N_t &= T_{21} P_{t|t} T_{21}^\top + Q_{22}, & L_t &= \bar{A}^\ell P_{t|t} T_{21}^\top N_t^{-1} \\ \hat{x}_{t+1|t} &= \bar{A}^\ell \hat{x}_{t|t} + Q_{21} Q_{22}^{-1} z_t + f_t^\ell + L_t (\lambda_{t+1} - f_t^n - T_{21} \hat{x}_{t|t}), & P_{t+1|t} &= \bar{A}^\ell P_{t|t} (\bar{A}^\ell)^\top + \bar{Q}^\ell - L_t N_t L_t^\top\end{aligned}$$

- (5) Set $t \leftarrow t + 1$ and go back to step (2).

Table A1. AR(1) least squares estimates for Nelson-Siegel-Svensson factors from European Central Bank. L_t is the level, S_t the slope, $C_{1,t}$, $C_{2,t}$ curvature factors with decay parameters $\lambda_{1,t}$, $\lambda_{2,t}$, respectively. (a) for factors estimated from AAA bonds and (c) for factors estimated from all bonds. μ is the sample mean, ρ is the AR(1) coefficient of the demeaned factor series (without an intercept), σ is the standard error of regression and T is the sample size. Standard errors in parentheses. The daily sample is 6 September 2004 to 21 February 2019.

	μ	ρ	σ	R^2	T
$L_t^{(a)}$	2.634 (0.025)	0.995 (0.002)	0.152	0.990	3701
$S_t^{(a)}$	−1.789 (0.015)	0.981 (0.003)	0.182	0.963	3701
$C_{1,t}^{(a)}$	11.663 (0.166)	0.994 (0.002)	1.064	0.989	3701
$C_{2,t}^{(a)}$	−11.339 (0.166)	0.995 (0.002)	1.027	0.990	3701
$\log \lambda_{1,t}^{(a)}$	0.966 (0.016)	0.994 (0.002)	0.107	0.987	3701
$\log \lambda_{2,t}^{(a)}$	0.831 (0.011)	0.994 (0.002)	0.071	0.988	3701
$L_t^{(c)}$	3.617 (0.024)	0.988 (0.003)	0.223	0.977	3701
$S_t^{(c)}$	−2.698 (0.027)	0.989 (0.002)	0.247	0.978	3701
$C_{1,t}^{(c)}$	16.070 (0.560)	0.980 (0.003)	6.800	0.960	3701
$C_{2,t}^{(c)}$	−17.210 (0.573)	0.981 (0.003)	6.794	0.962	3701
$\log \lambda_{1,t}^{(c)}$	0.774 (0.015)	0.992 (0.002)	0.111	0.984	3701
$\log \lambda_{2,t}^{(c)}$	0.565 (0.011)	0.989 (0.002)	0.105	0.977	3701

Table A2. Maximum likelihood estimates of four dynamic Nelson-Siegel-Svensson type specifications for zero yield curve data from the European Central Bank (AAA bonds). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The likelihood functions were evaluated using the extended Kalman filter. σ_τ are the estimated measurement equation innovation variances (assumed to be diagonal), $\rho(\cdot)$ are the estimated AR(1) parameters of the state series, $E[\cdot]$ are the estimated unconditional mean of the state series, $\sigma(\cdot)$ are the estimated standard deviations of the state equation innovations. QML standard errors in parentheses. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations). The measurement consists of 13 yields with maturities $\tau = 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25$ (years).

	<i>DNSS</i>		<i>DNS</i>		<i>NSS</i>		<i>NS</i>	
σ_1	0.038	(0.007)	0.063	(0.001)	0.076	(0.003)	0.106	(0.003)
$\sigma_{1.5}$	0.011	(0.001)	0.020	(0.001)	0.024	(0.001)	0.037	(0.001)
σ_2	0.002	(0.000)	0.004	(0.001)	0.000	(0.000)	0.000	(0.000)
σ_3	0.004	(0.000)	0.008	(0.000)	0.011	(0.000)	0.024	(0.000)

Table A2. Cont.

	DNSS		DNS		NSS		NS	
σ_4	0.002	(0.000)	0.004	(0.000)	0.007	(0.000)	0.022	(0.000)
σ_5	0.001	(0.000)	0.003	(0.000)	0.002	(0.001)	0.014	(0.000)
σ_7	0.000	(0.000)	0.001	(0.000)	0.005	(0.000)	0.000	(0.000)
σ_8	0.000	(0.000)	0.001	(0.000)	0.004	(0.000)	0.002	(0.000)
σ_9	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_{10}	0.000	(0.000)	0.003	(0.000)	0.005	(0.000)	0.005	(0.000)
σ_{15}	0.003	(0.000)	0.057	(0.001)	0.020	(0.001)	0.069	(0.001)
σ_{20}	0.007	(0.002)	0.168	(0.002)	0.017	(0.004)	0.176	(0.002)
σ_{25}	0.018	(0.005)	0.305	(0.004)	0.069	(0.008)	0.303	(0.003)
$\rho(L)$	0.998	(0.020)	0.999	(0.001)	1.000	(0.000)	0.999	(0.000)
$\rho(S)$	0.998	(0.004)	0.999	(0.001)	0.999	(0.001)	0.999	(0.001)
$\rho(C_1)$	1.000	(0.000)	0.994	(0.002)	0.998	(0.001)	0.995	(0.002)
$\rho(C_2)$	0.985	(0.002)			0.996	(0.002)		
$\rho(\lambda_1)$	1.000	(0.000)	0.996	(0.001)				
$\rho(\lambda_2)$	1.000	(0.000)						
$E[L]$	2.998	(1.945)	3.066	(0.666)	3.025	(1.715)	3.515	(1.223)
$E[S]$	0.917	(0.390)	−0.760	(1.658)	−2.115	(0.668)	−2.648	(0.728)
$E[C_1]$	−3.115	(0.091)	−4.345	(0.168)	3.561	(1.247)	−3.057	(0.519)
$E[C_2]$	0.000	(0.000)			−2.679	(0.813)		
$E[\lambda_1]$	0.001	(0.000)	0.000	(0.000)	0.162	(0.002)	0.415	(0.002)
$E[\lambda_2]$	0.001	(0.000)			0.340	(0.002)		
$\sigma(L)$	0.061	(0.006)	0.049	(0.001)	0.040	(0.002)	0.056	(0.001)
$\sigma(S)$	0.081	(0.010)	0.054	(0.001)	0.042	(0.002)	0.067	(0.002)
$\sigma(C_1)$	0.155	(0.003)	0.150	(0.004)	0.177	(0.005)	0.159	(0.005)
$\sigma(C_2)$	0.282	(0.022)			0.189	(0.009)		
$\sigma(\lambda_1)$	0.029	(0.002)	0.033	(0.002)				
$\sigma(\lambda_2)$	0.035	(0.003)						

Table A3. Maximum likelihood estimates of four dynamic Nelson-Siegel-Svensson type specifications for zero yield curve data from the European Central Bank (AAA bonds). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The likelihood functions were evaluated using the extended Kalman filter. σ_τ are the estimated measurement equation innovation variances (assumed to be diagonal), $\rho(\cdot)$ are the estimated AR(1) parameters of the state series, $E[\cdot]$ are the estimated unconditional mean of the state series, $\sigma(\cdot)$ are the estimated standard deviations of the state equation innovations. QML standard errors in parentheses. The daily sample is from 6 September 2004 to 21 February 2019 (3701 observations). The measurement consists of 9 yields with maturities $\tau = 1/12, 0.5, 1, 2, 3, 5, 7, 10, 20$ (years).

	DNSS		DNS		NSS		NS	
$\sigma_{1/12}$	0.122	(0.026)	0.278	(0.012)	0.274	(0.015)	0.267	(0.006)
$\sigma_{0.5}$	0.000	(0.000)	0.068	(0.001)	0.052	(0.002)	0.082	(0.002)
σ_1	0.003	(0.001)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_2	0.003	(0.001)	0.029	(0.001)	0.008	(0.001)	0.043	(0.001)
σ_3	0.000	(0.000)	0.020	(0.001)	0.002	(0.001)	0.030	(0.001)
σ_5	0.001	(0.000)	0.005	(0.001)	0.000	(0.000)	0.000	(0.000)
σ_7	0.000	(0.001)	0.004	(0.001)	0.005	(0.000)	0.007	(0.001)
σ_{10}	0.004	(0.000)	0.011	(0.001)	0.012	(0.002)	0.030	(0.001)
σ_{20}	0.142	(0.002)	0.168	(0.002)	0.163	(0.003)	0.215	(0.004)
$\rho(L)$	1.000	(0.000)	0.999	(0.000)	1.000	(0.000)	1.000	(0.000)
$\rho(S)$	0.998	(0.002)	1.000	(0.000)	0.999	(0.000)	1.000	(0.000)

Table A3. Cont.

	DNSS		DNS		NSS		NS	
$\rho(C_1)$	1.000	(0.001)	0.997	(0.001)	0.997	(0.002)	0.996	(0.001)
$\rho(C_2)$	0.997	(0.002)			0.992	(0.006)		
$\rho(\lambda_1)$	0.976	(0.005)	0.998	(0.001)				
$\rho(\lambda_2)$	0.998	(0.001)						
$E[L]$	2.588	(1.561)	2.843	(0.820)	3.390	(1.568)	3.337	(1.769)
$E[S]$	−2.808	(0.848)	−2.410	(0.099)	−2.739	(0.702)	−2.632	(0.812)
$E[C_1]$	3.406	(2.469)	−2.751	(0.214)	−3.675	(0.716)	−2.828	(0.560)
$E[C_2]$	−4.948	(0.601)			0.910	(0.328)		
$E[\lambda_1]$	0.276	(0.053)	0.000	(0.000)	0.451	(0.002)	0.410	(0.002)
$E[\lambda_2]$	0.000	(0.000)			0.794	(0.008)		
$\sigma(L)$	0.042	(0.001)	0.024	(0.001)	0.040	(0.001)	0.033	(0.001)
$\sigma(S)$	0.061	(0.005)	0.028	(0.002)	0.045	(0.002)	0.041	(0.001)
$\sigma(C_1)$	0.103	(0.006)	0.119	(0.003)	0.156	(0.005)	0.134	(0.003)
$\sigma(C_2)$	0.144	(0.006)			0.170	(0.007)		
$\sigma(\lambda_1)$	0.037	(0.003)	0.025	(0.001)				
$\sigma(\lambda_2)$	0.021	(0.001)						

Table A4. Maximum likelihood estimates of four dynamic Nelson-Siegel-Svensson type specifications for zero yield curve data from the European Central Bank (all ratings). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The likelihood functions were evaluated using the extended Kalman filter. σ_τ are the estimated measurement equation innovation variances (assumed to be diagonal), $\rho(\cdot)$ are the estimated AR(1) parameters of the state series, $E[\cdot]$ are the estimated unconditional mean of the state series, $\sigma(\cdot)$ are the estimated standard deviations of the state equation innovations. QML standard errors in parentheses. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations). The measurement consists of 13 yields with maturities $\tau = 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25$ (years).

	DNSS		DNS		NSS		NS	
σ_1	0.004	(0.001)	0.245	(0.011)	0.116	(0.003)	0.126	(0.003)
$\sigma_{1.5}$	0.000	(0.000)	0.100	(0.006)	0.037	(0.001)	0.044	(0.001)
σ_2	0.001	(0.000)	0.038	(0.002)	0.014	(0.001)	0.018	(0.002)
σ_3	0.000	(0.000)	0.009	(0.001)	0.028	(0.001)	0.041	(0.001)
σ_4	0.000	(0.000)	0.004	(0.000)	0.025	(0.000)	0.039	(0.001)
σ_5	0.000	(0.000)	0.003	(0.000)	0.014	(0.000)	0.025	(0.000)
σ_7	0.000	(0.000)	0.001	(0.000)	0.000	(0.000)	0.003	(0.000)
σ_8	0.000	(0.000)	0.000	(0.000)	0.001	(0.000)	0.002	(0.000)
σ_9	0.001	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_{10}	0.002	(0.000)	0.001	(0.000)	0.002	(0.000)	0.005	(0.000)
σ_{15}	0.026	(0.001)	0.031	(0.001)	0.006	(0.001)	0.054	(0.001)
σ_{20}	0.089	(0.002)	0.099	(0.002)	0.046	(0.001)	0.121	(0.002)
σ_{25}	0.178	(0.004)	0.192	(0.004)	0.117	(0.002)	0.206	(0.004)
$\rho(L)$	0.998	(0.001)	0.998	(0.001)	0.999	(0.001)	0.999	(0.001)
$\rho(S)$	1.000	(0.001)	0.998	(0.004)	0.998	(0.001)	0.998	(0.001)
$\rho(C_1)$	0.999	(0.000)	0.993	(0.005)	0.996	(0.002)	0.993	(0.002)
$\rho(C_2)$	0.997	(0.001)			0.993	(0.002)		
$\rho(\lambda_1)$	0.817	(0.022)	0.991	(0.002)				
$\rho(\lambda_2)$	0.989	(0.003)						
$E[L]$	3.960	(0.400)	3.628	(0.499)	4.154	(0.611)	4.023	(0.946)
$E[S]$	−6.188	(3.230)	−0.427	(1.488)	−2.951	(0.499)	−2.851	(0.638)
$E[C_1]$	32.304	(8.643)	−5.720	(0.976)	1.422	(0.853)	−3.293	(0.441)

Table A4. Cont.

	DNSS		DNS		NSS		NS	
$E[C_2]$	−13.254	(1.995)			−2.715	(0.685)		
$E[\lambda_1]$	0.118	(0.019)	0.000	(0.000)	0.241	(0.005)	0.451	(0.002)
$E[\lambda_2]$	0.000	(0.000)			0.417	(0.003)		
$\sigma(L)$	0.060	(0.002)	0.064	(0.002)	0.059	(0.002)	0.054	(0.001)
$\sigma(S)$	0.105	(0.006)	0.084	(0.003)	0.072	(0.003)	0.085	(0.003)
$\sigma(C_1)$	0.230	(0.012)	0.202	(0.008)	0.217	(0.008)	0.186	(0.007)
$\sigma(C_2)$	0.155	(0.008)			0.290	(0.009)		
$\sigma(\lambda_1)$	0.059	(0.006)	0.041	(0.002)				
$\sigma(\lambda_2)$	0.053	(0.003)						

Table A5. Maximum likelihood estimates of four dynamic Nelson-Siegel-Svensson type specifications for zero yield curve data from the European Central Bank (all ratings). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The likelihood functions were evaluated using the extended Kalman filter. σ_τ are the estimated measurement equation innovation variances (assumed to be diagonal), $\rho(\cdot)$ are the estimated AR(1) parameters of the state series, $E[\cdot]$ are the estimated unconditional mean of the state series, $\sigma(\cdot)$ are the estimated standard deviations of the state equation innovations. QML standard errors in parentheses. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations). The measurement consists of 9 yields with maturities $\tau = 1/12, 0.5, 1, 2, 3, 5, 7, 10, 20$ (years).

	DNSS		DNS		NSS		NS	
$\sigma_{1/12}$	0.061	(0.013)	0.528	(0.009)	0.243	(0.007)	0.496	(0.009)
$\sigma_{0.5}$	0.006	(0.001)	0.161	(0.003)	0.000	(0.000)	0.151	(0.003)
σ_1	0.004	(0.000)	0.000	(0.000)	0.039	(0.001)	0.011	(0.002)
σ_2	0.002	(0.000)	0.049	(0.001)	0.014	(0.001)	0.056	(0.001)
σ_3	0.001	(0.000)	0.033	(0.001)	0.022	(0.001)	0.042	(0.001)
σ_5	0.001	(0.000)	0.022	(0.000)	0.001	(0.003)	0.009	(0.001)
σ_7	0.001	(0.000)	0.006	(0.001)	0.016	(0.000)	0.020	(0.000)
σ_{10}	0.004	(0.000)	0.031	(0.001)	0.016	(0.001)	0.014	(0.001)
σ_{20}	0.102	(0.002)	0.025	(0.002)	0.129	(0.002)	0.154	(0.002)
$\rho(L)$	0.998	(0.001)	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
$\rho(S)$	0.998	(0.001)	0.999	(0.001)	0.999	(0.000)	0.999	(0.000)
$\rho(C_1)$	0.990	(0.002)	0.997	(0.001)	0.996	(0.001)	0.996	(0.001)
$\rho(C_2)$	1.000	(0.000)			0.989	(0.006)		
$\rho(\lambda_1)$	0.996	(0.001)	1.000	(0.000)				
$\rho(\lambda_2)$	0.931	(0.015)						
$E[L]$	4.214	(0.480)	4.191	(0.776)	3.760	(1.348)	3.871	(1.325)
$E[S]$	−3.486	(0.116)	−2.776	(0.105)	−3.152	(0.637)	−3.042	(0.698)
$E[C_1]$	−4.892	(0.543)	−3.219	(0.167)	−3.552	(0.558)	−2.714	(0.567)
$E[C_2]$	5.434	(3.799)			1.425	(0.263)		
$E[\lambda_1]$	0.000	(0.000)	0.000	(0.000)	0.496	(0.003)	0.409	(0.002)
$E[\lambda_2]$	0.278	(0.002)			1.227	(0.016)		
$\sigma(L)$	0.054	(0.002)	0.033	(0.001)	0.035	(0.001)	0.035	(0.001)
$\sigma(S)$	0.076	(0.005)	0.048	(0.002)	0.039	(0.001)	0.054	(0.002)
$\sigma(C_1)$	0.194	(0.011)	0.107	(0.003)	0.162	(0.004)	0.136	(0.004)
$\sigma(C_2)$	0.144	(0.014)			0.171	(0.008)		
$\sigma(\lambda_1)$	0.037	(0.002)	0.027	(0.001)				
$\sigma(\lambda_2)$	0.055	(0.008)						

Table A6. Maximum likelihood estimates of four dynamic Nelson-Siegel-Svensson type specifications for zero yield curve data from the Bank of Canada. *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The likelihood functions were evaluated using the extended Kalman filter. σ_τ are the estimated measurement equation innovation variances (assumed to be diagonal), $\rho(\cdot)$ are the estimated AR(1) parameters of the state series, $E[\cdot]$ are the estimated unconditional mean of the state series, $\sigma(\cdot)$ are the estimated standard deviations of the state equation innovations. QML standard errors in parentheses. The daily sample is 4 January 2000 to 31 December 2018 (4,719 observations). The measurement consists of 16 yields with maturities $\tau = 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30$ (years).

	<i>DNSS</i>		<i>DNS</i>		<i>NSS</i>		<i>NS</i>	
$\sigma_{0.25}$	0.000	(0.000)	0.001	(0.000)	0.001	(0.000)	0.002	(0.000)
$\sigma_{0.5}$	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.001	(0.000)
$\sigma_{0.75}$	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.001	(0.000)
σ_1	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
$\sigma_{1.5}$	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_2	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_3	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_4	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_5	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_7	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_8	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_9	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_{10}	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_{15}	0.001	(0.000)	0.001	(0.000)	0.002	(0.000)	0.002	(0.000)
σ_{20}	0.002	(0.000)	0.001	(0.000)	0.002	(0.000)	0.002	(0.000)
σ_{30}	0.004	(0.000)	0.004	(0.000)	0.004	(0.000)	0.003	(0.000)
$\rho(L)$	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
$\rho(S)$	1.000	(0.000)	1.000	(0.000)	0.999	(0.001)	0.999	(0.000)
$\rho(C_1)$	1.000	(0.000)	0.988	(0.001)	0.996	(0.002)	0.998	(0.001)
$\rho(C_2)$	0.999	(0.000)			0.999	(0.001)		
$\rho(\lambda_1)$	0.959	(0.004)	0.994	(0.001)				
$\rho(\lambda_2)$	0.996	(0.002)						
$E[L]$	0.125	(0.302)	0.053	(0.012)	0.043	(0.013)	0.042	(0.012)
$E[S]$	0.429	(0.015)	0.062	(0.021)	−0.020	(0.006)	−0.017	(0.007)
$E[C_1]$	−0.083	(0.023)	−0.028	(0.002)	−0.010	(0.005)	−0.011	(0.009)
$E[C_2]$	−0.093	(0.017)			−0.025	(0.012)		
$E[\lambda_1]$	0.632	(0.000)	0.000	(0.000)	1.466	(0.009)	0.419	(0.003)
$E[\lambda_2]$	0.001	(0.013)			0.514	(0.003)		
$\sigma(L)$	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
$\sigma(S)$	0.001	(0.000)	0.000	(0.000)	0.001	(0.000)	0.001	(0.000)
$\sigma(C_1)$	0.001	(0.000)	0.001	(0.000)	0.001	(0.000)	0.001	(0.000)
$\sigma(C_2)$	0.001	(0.000)			0.001	(0.000)		
$\sigma(\lambda_1)$	0.021	(0.002)	0.029	(0.001)				
$\sigma(\lambda_2)$	0.027	(0.001)						

Table A7. Maximum likelihood estimates of four dynamic Nelson-Siegel-Svensson type specifications for zero yield curve data from the Bank of England. *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The likelihood functions were evaluated using the extended Kalman filter. σ_τ are the estimated measurement equation innovation variances (assumed to be diagonal), $\rho(\cdot)$ are the estimated AR(1) parameters of the state series, $E[\cdot]$ are the estimated unconditional mean of the state series, $\sigma(\cdot)$ are the estimated standard deviations of the state equation innovations. QML standard errors in parentheses. The daily sample is 4 January 2000 to 31 December 2018 (4,801 observations). The measurement consists of 13 yields with maturities $\tau = 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25$ (years).

	<i>DNSS</i>		<i>DNS</i>		<i>NSS</i>		<i>NS</i>	
σ_1	0.081	(0.003)	0.079	(0.001)	0.095	(0.001)	0.114	(0.002)
$\sigma_{1.5}$	0.020	(0.001)	0.030	(0.001)	0.023	(0.000)	0.040	(0.001)
σ_2	0.003	(0.002)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
σ_3	0.004	(0.001)	0.027	(0.001)	0.008	(0.000)	0.031	(0.001)
σ_4	0.002	(0.000)	0.027	(0.001)	0.003	(0.000)	0.031	(0.001)
σ_5	0.002	(0.000)	0.017	(0.001)	0.006	(0.000)	0.020	(0.001)
σ_7	0.001	(0.000)	0.001	(0.000)	0.005	(0.000)	0.000	(0.000)
σ_8	0.001	(0.000)	0.002	(0.000)	0.000	(0.000)	0.003	(0.000)
σ_9	0.000	(0.000)	0.000	(0.000)	0.007	(0.000)	0.000	(0.000)
σ_{10}	0.002	(0.000)	0.005	(0.000)	0.014	(0.000)	0.007	(0.000)
σ_{15}	0.021	(0.001)	0.058	(0.001)	0.034	(0.001)	0.088	(0.001)
σ_{20}	0.023	(0.004)	0.130	(0.002)	0.015	(0.002)	0.187	(0.002)
σ_{25}	0.059	(0.002)	0.218	(0.002)	0.087	(0.001)	0.283	(0.003)
$\rho(L)$	1.000	(0.001)	0.996	(0.001)	0.998	(0.002)	0.999	(0.001)
$\rho(S)$	0.999	(0.001)	0.997	(0.001)	1.000	(0.000)	0.999	(0.000)
$\rho(C_1)$	1.000	(0.000)	0.999	(0.000)	0.997	(0.001)	0.999	(0.001)
$\rho(C_2)$	1.000	(0.000)			0.998	(0.001)		
$\rho(\lambda_1)$	0.973	(0.012)	0.998	(0.000)				
$\rho(\lambda_2)$	0.998	(0.001)						
$E[L]$	4.255	(149.500)	4.417	(0.174)	3.453	(0.486)	4.016	(0.450)
$E[S]$	7.922	(5.027)	−2.280	(0.134)	0.394	(4.183)	−0.962	(1.868)
$E[C_1]$	−39.922	(3.827)	4.444	(0.677)	2.355	(1.028)	−0.507	(2.401)
$E[C_2]$	10.262	(1.698)			0.525	(1.207)		
$E[\lambda_1]$	0.037	(0.005)	0.000	(0.000)	0.137	(0.001)	0.396	(0.002)
$E[\lambda_2]$	0.001	(0.006)			2.333	(0.009)		
$\sigma(L)$	0.056	(0.003)	0.054	(0.001)	0.060	(0.003)	0.059	(0.001)
$\sigma(S)$	0.056	(0.003)	0.067	(0.002)	0.075	(0.002)	0.077	(0.002)
$\sigma(C_1)$	0.164	(0.021)	0.123	(0.003)	0.216	(0.008)	0.120	(0.003)
$\sigma(C_2)$	0.219	(0.022)			0.176	(0.006)		
$\sigma(\lambda_1)$	0.029	(0.013)	0.025	(0.001)				
$\sigma(\lambda_2)$	0.016	(0.001)						

Table A8. Maximum likelihood estimates of four dynamic Nelson-Siegel-Svensson type specifications for zero yield curve data from the U.S. Department of Treasury. *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The likelihood functions were evaluated using the extended Kalman filter. σ_τ are the estimated measurement equation innovation variances (assumed to be diagonal), $\rho(\cdot)$ are the estimated AR(1) parameters of the state series, $E[\cdot]$ are the estimated unconditional mean of the state series, $\sigma(\cdot)$ are the estimated standard deviations of the state equation innovations. QML standard errors in parentheses. The daily sample is 31 July 2001 to 31 December 2018 (4,357 observations). The measurement consists of 9 yields with maturities $\tau = 1/12, 0.5, 1, 2, 3, 5, 7, 10, 20$ (years).

	<i>DNSS</i>		<i>DNS</i>		<i>NSS</i>		<i>NS</i>	
$\sigma_{1/12}$	0.001	(0.000)	0.240	(0.005)	0.241	(0.005)	0.251	(0.005)
$\sigma_{0.5}$	0.054	(0.001)	0.000	(0.000)	0.000	(0.000)	0.007	(0.005)
σ_1	0.014	(0.002)	0.044	(0.001)	0.055	(0.001)	0.064	(0.001)
σ_2	0.029	(0.001)	0.048	(0.003)	0.045	(0.001)	0.056	(0.001)
σ_3	0.015	(0.001)	0.035	(0.003)	0.020	(0.002)	0.000	(0.000)
σ_5	0.038	(0.001)	0.030	(0.001)	0.040	(0.001)	0.071	(0.001)
σ_7	0.001	(0.003)	0.042	(0.001)	0.000	(0.000)	0.076	(0.002)
σ_{10}	0.077	(0.001)	0.083	(0.003)	0.091	(0.002)	0.036	(0.008)
σ_{20}	0.014	(0.002)	0.045	(0.014)	0.023	(0.004)	0.159	(0.002)
$\rho(L)$	1.000	(0.000)	0.999	(0.000)	1.000	(0.000)	0.999	(0.000)
$\rho(S)$	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
$\rho(C_1)$	0.995	(0.001)	0.996	(0.001)	0.998	(0.001)	0.994	(0.001)
$\rho(C_2)$	0.999	(0.001)			0.994	(0.002)		
$\rho(\lambda_1)$	0.995	(0.001)	0.998	(0.001)				
$\rho(\lambda_2)$	0.998	(0.001)						
$E[L]$	9.630	(5.932)	4.492	(0.765)	4.524	(0.957)	4.433	(0.934)
$E[S]$	0.688	(0.549)	−2.343	(0.181)	−2.315	(0.795)	−2.233	(0.860)
$E[C_1]$	−0.814	(0.483)	−3.273	(0.225)	−2.456	(0.804)	−2.475	(0.511)
$E[C_2]$	−5.540	(0.391)			−0.051	(0.502)		
$E[\lambda_1]$	0.805	(0.079)	0.000	(0.000)	0.512	(0.006)	0.506	(0.003)
$E[\lambda_2]$	0.000	(0.000)			0.218	(0.013)		
$\sigma(L)$	0.073	(0.014)	0.037	(0.003)	0.039	(0.004)	0.042	(0.002)
$\sigma(S)$	0.096	(0.007)	0.041	(0.003)	0.045	(0.003)	0.050	(0.002)
$\sigma(C_1)$	0.107	(0.005)	0.136	(0.013)	0.150	(0.005)	0.213	(0.013)
$\sigma(C_2)$	0.142	(0.008)			0.227	(0.023)		
$\sigma(\lambda_1)$	0.022	(0.001)	0.038	(0.003)				
$\sigma(\lambda_2)$	0.023	(0.001)						

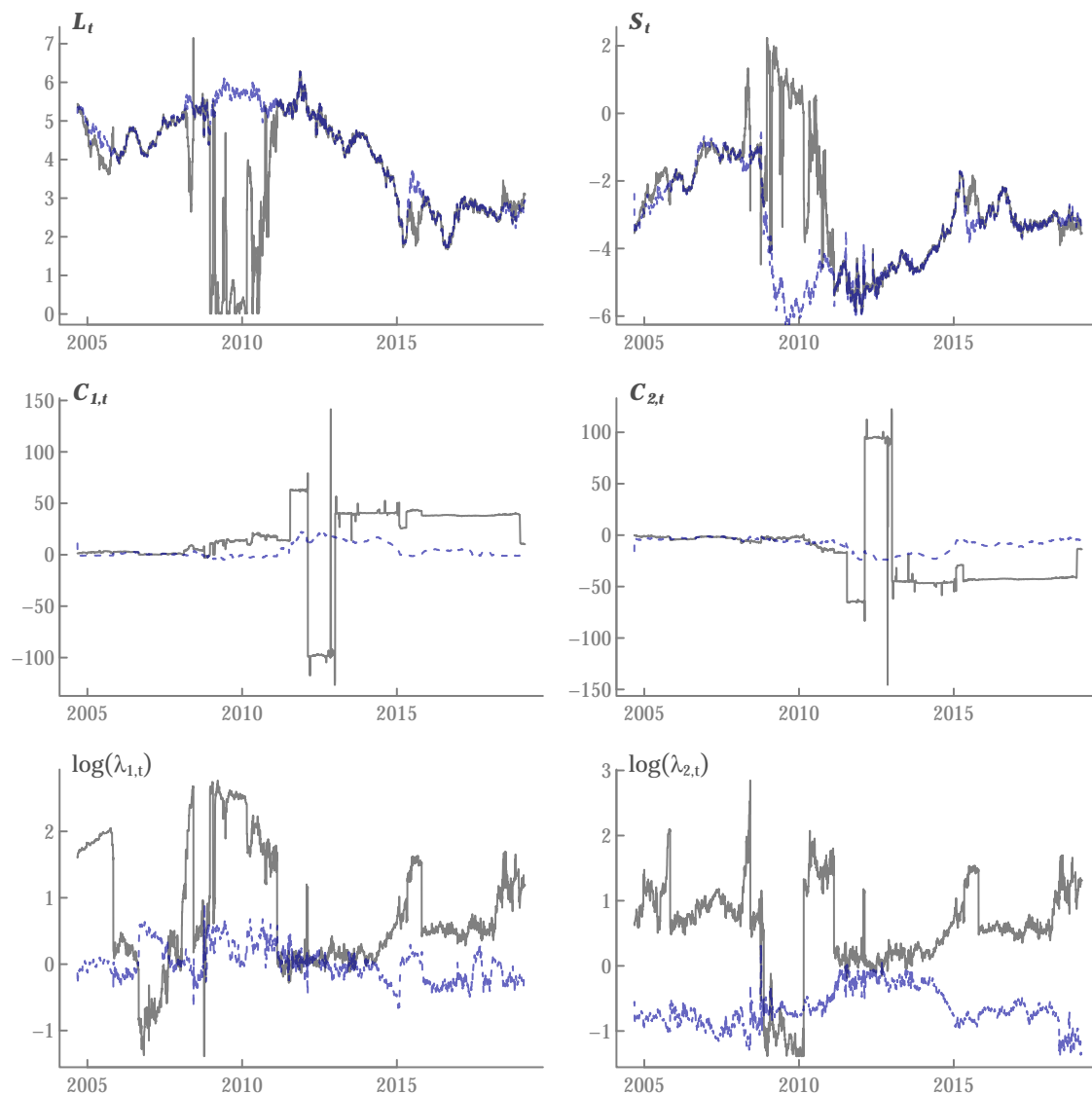


Figure A1. Nelson-Siegel-Svensson factors (all ratings) from the European Central Bank. L_t is the level, S_t the slope, $C_{1,t}$, $C_{2,t}$ curvature factors with decay parameters $\lambda_{1,t}$, $\lambda_{2,t}$, respectively. The gray solid lines are the actual factors and the blue dashed lines are filtered estimates from the extended Kalman filter. The parameters of the state space model were estimated using 13 zero yield curves with maturities $\tau = 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25$ (years). The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

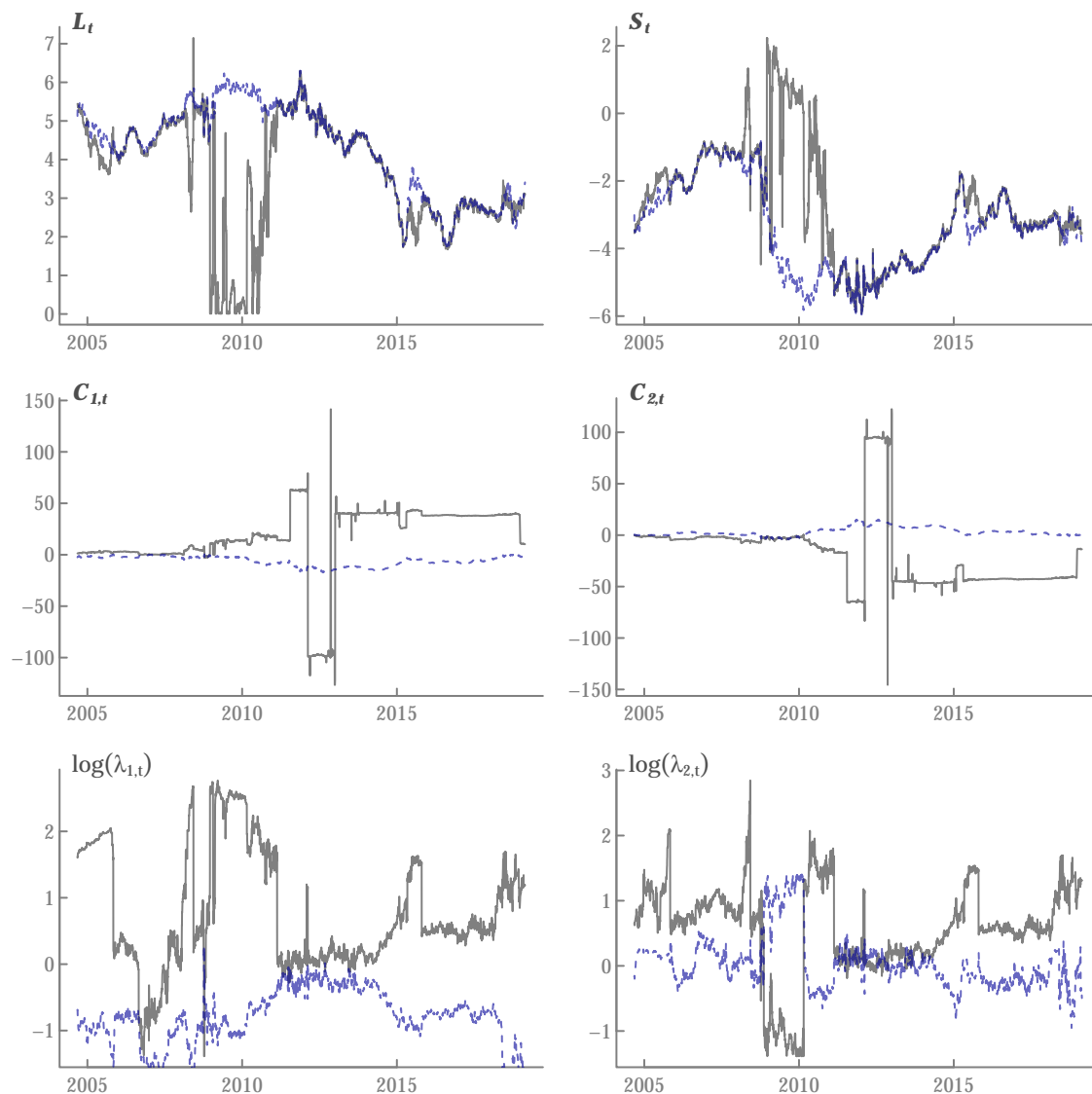


Figure A2. Nelson-Siegel-Svensson factors (all ratings) from the European Central Bank. L_t is the level, S_t the slope, $C_{1,t}$, $C_{2,t}$ curvature factors with decay parameters $\lambda_{1,t}$, $\lambda_{2,t}$, respectively. The gray solid lines are the actual factors and the blue dashed lines are filtered estimates from the extended Kalman filter. The parameters of the state space model were estimated using 9 zero yield curves with maturities $\tau = 1/12, 0.5, 1, 2, 3, 5, 7, 10, 20$ (years). The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

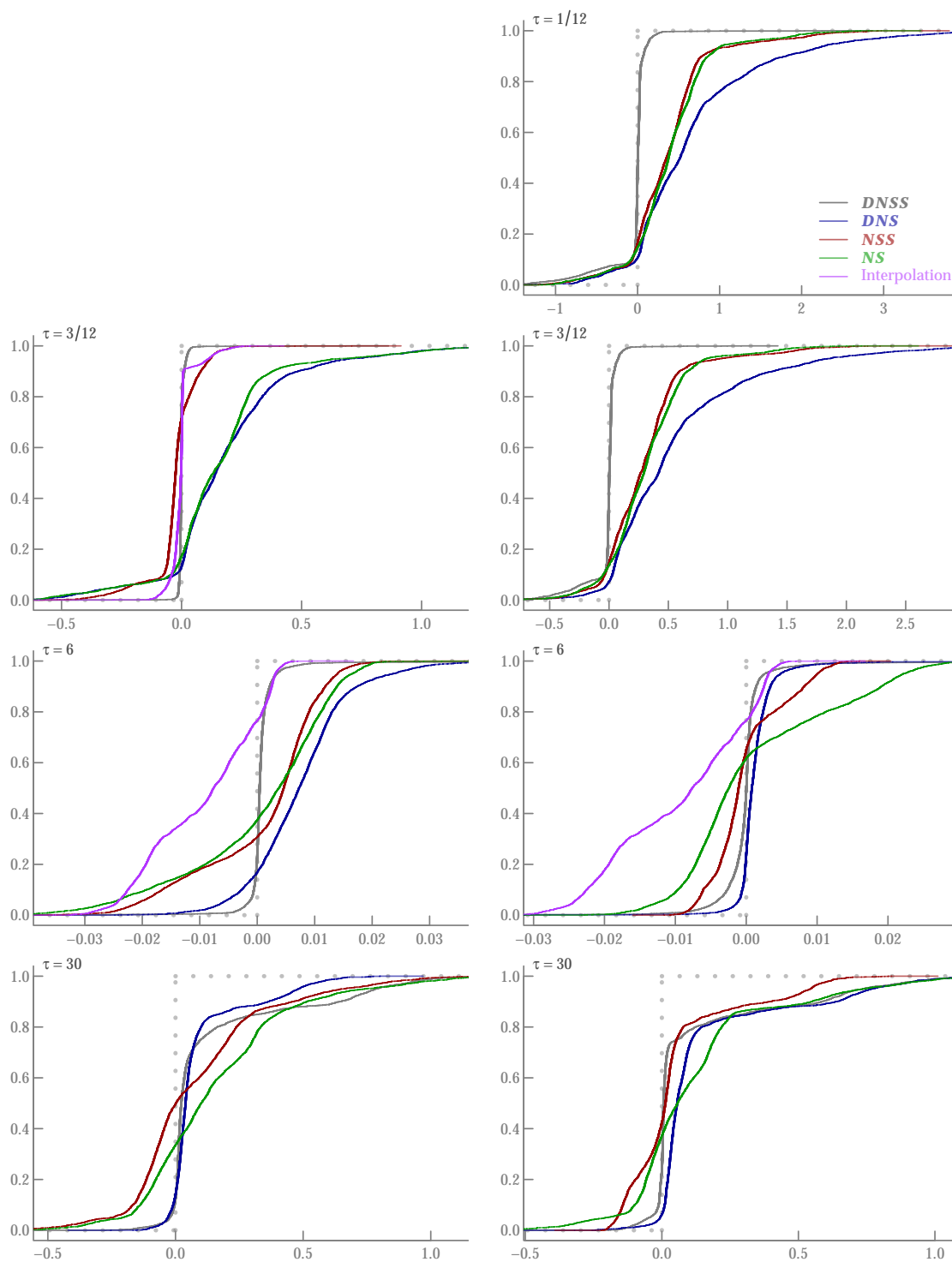


Figure A3. Empirical cdf of prediction errors. Predictions from European Central Bank (all ratings) estimates with maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ (left panel) and $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$ (right panel). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The distribution of a perfect prediction with no error would be a vertical line at $e_t = 0$ indicated as a dotted line. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

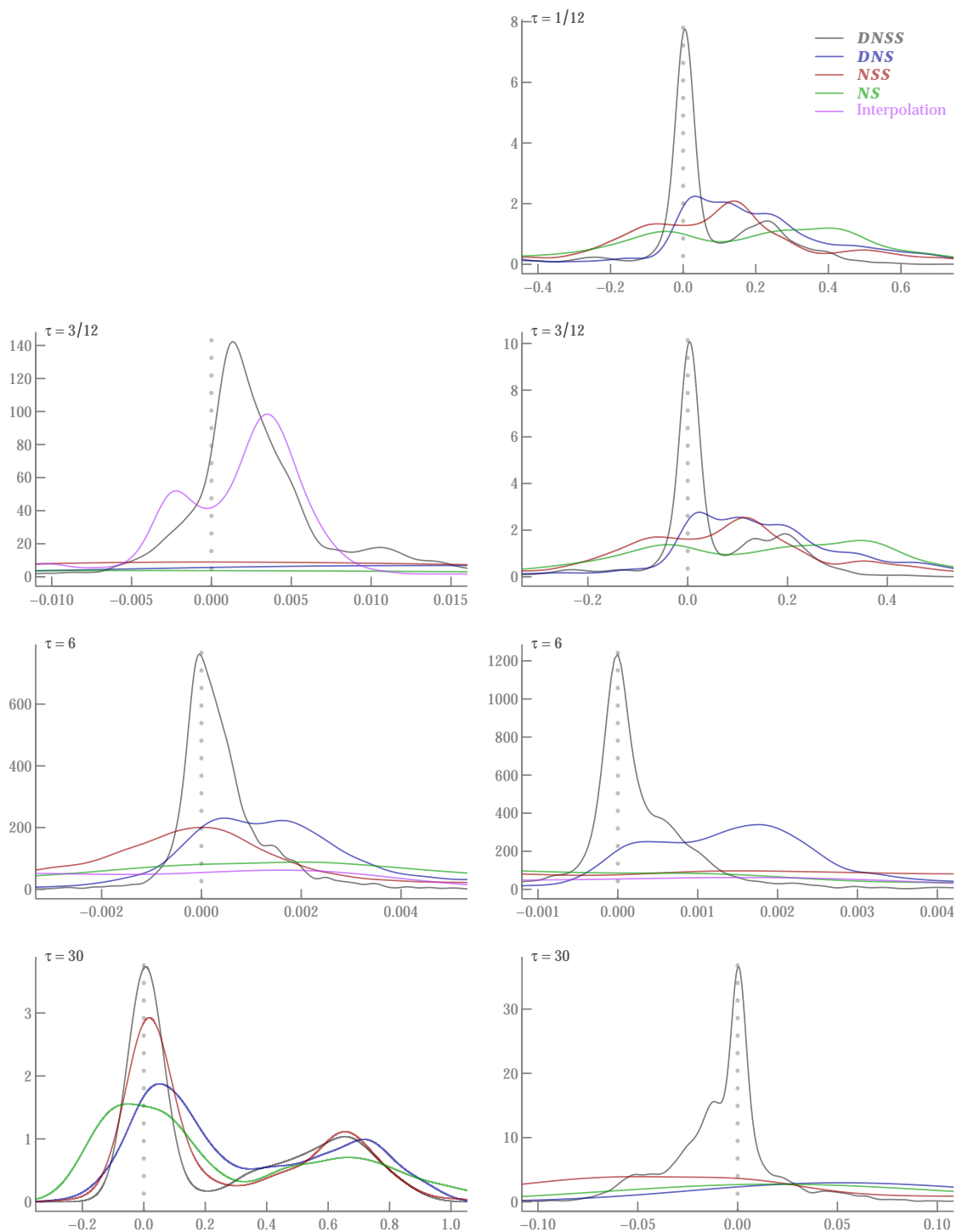


Figure A4. Kernel density estimates of prediction error distributions. Predictions from European Central Bank (AAA ratings) estimates with maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ (left panel) and $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$ (right panel). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

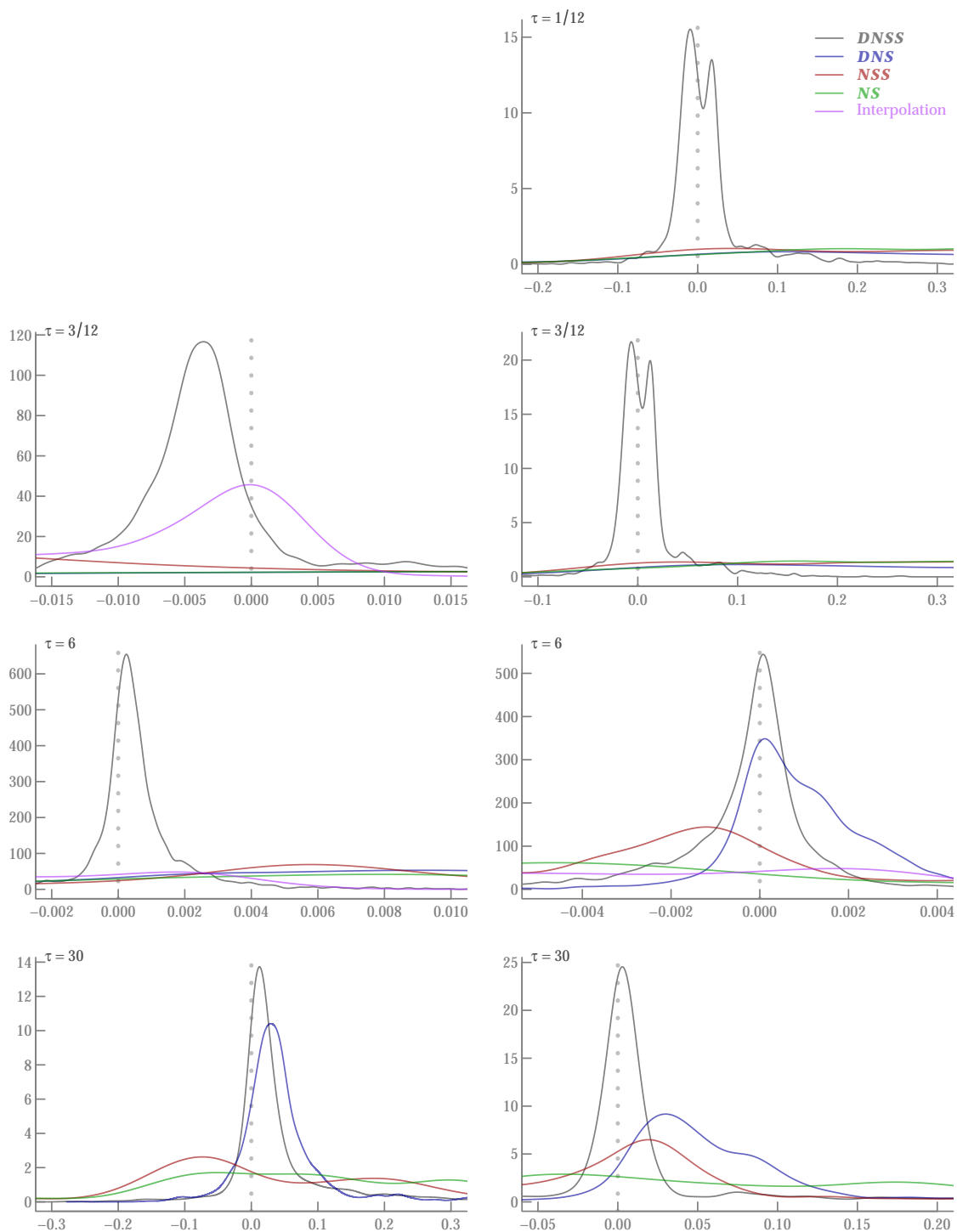


Figure A5. Kernel density estimates of prediction error distributions. Predictions from European Central Bank (all ratings) estimates with maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ (left panel) and $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$ (right panel). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The daily sample is 6 September 2004 to 21 February 2019 (3701 observations).

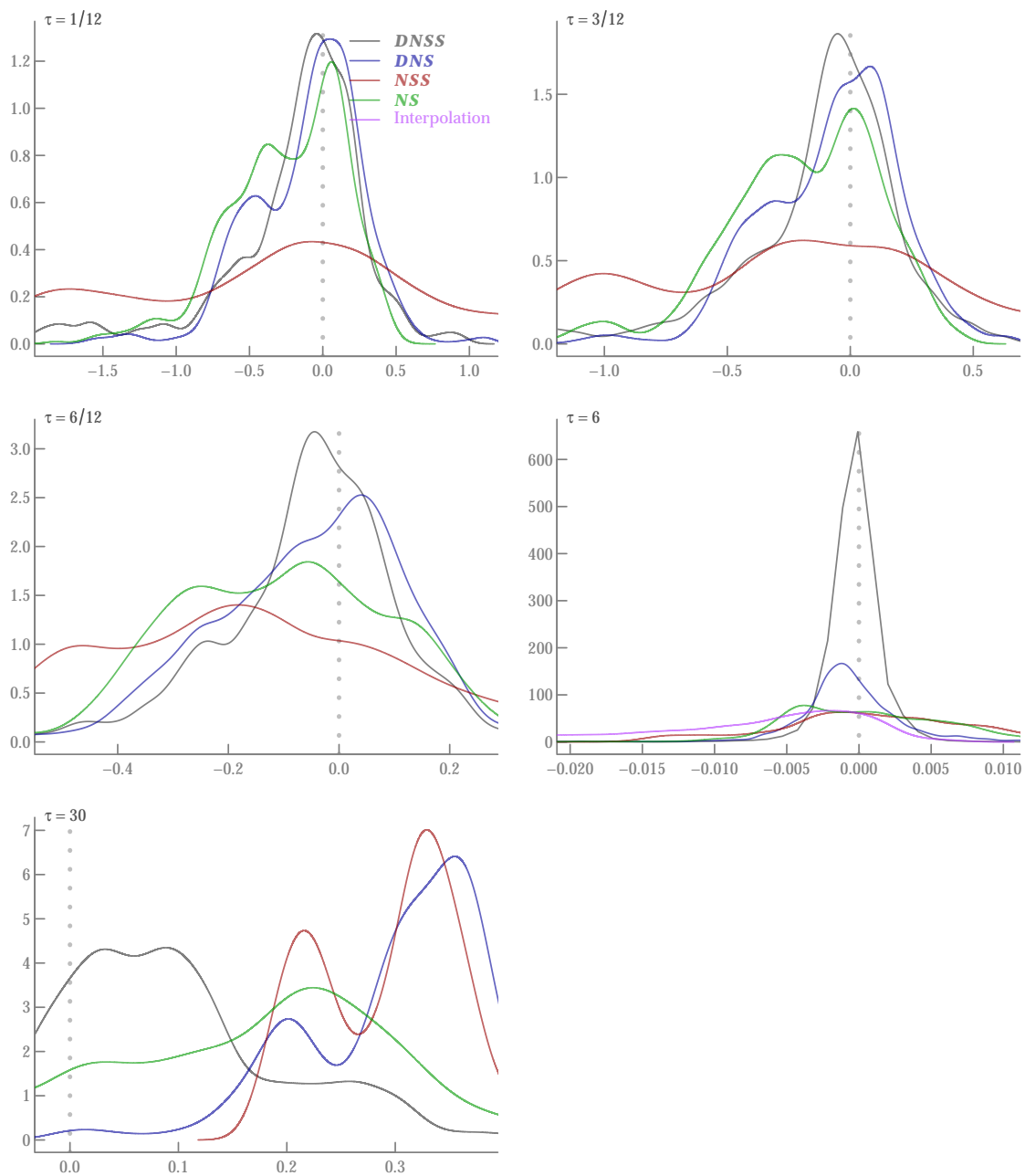


Figure A6. Kernel density estimates of prediction error distributions. Predictions from Bank of England data with maturities $\tau_{13} = (1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25)$. *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The daily sample is 4 January 2000 to 31 December 2018 (4801 observations some of which may be missing).

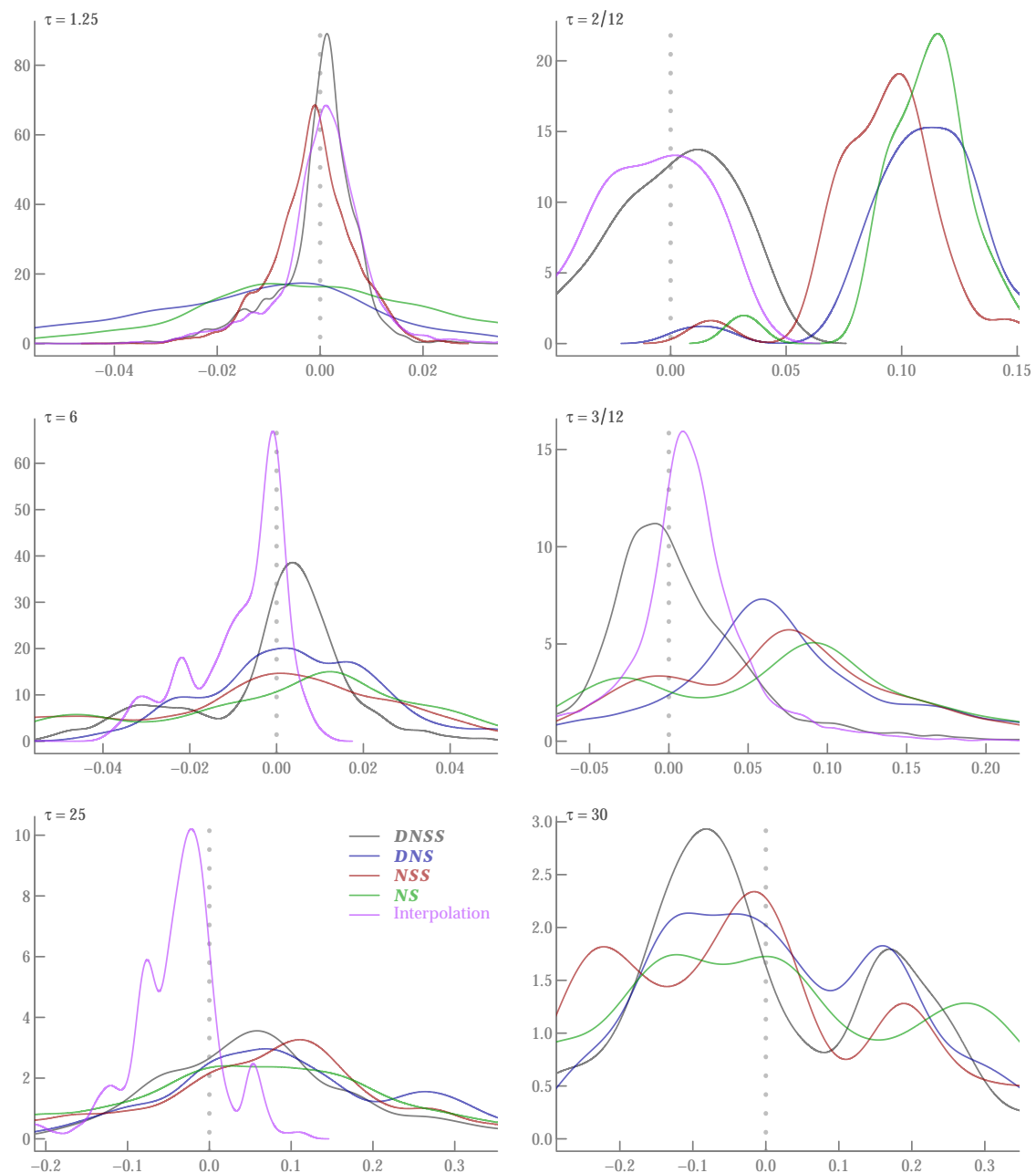


Figure A7. Kernel density estimates of prediction error distributions. Predictions from Bank of Canada yield data with maturities $\tau_{16} = (3/12, 6/12, 0.75, 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30)$ (left panel) and U.S. Department of Treasury with maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ (right panel). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The daily sample is 4 January 2000 to 31 December 2018 (4719 observations) for Bank of Canada (left panel) and 31 July 2001 to 31 December 2018 (4357 observations some of which may be missing) for U.S. Department of Treasury (right panel).

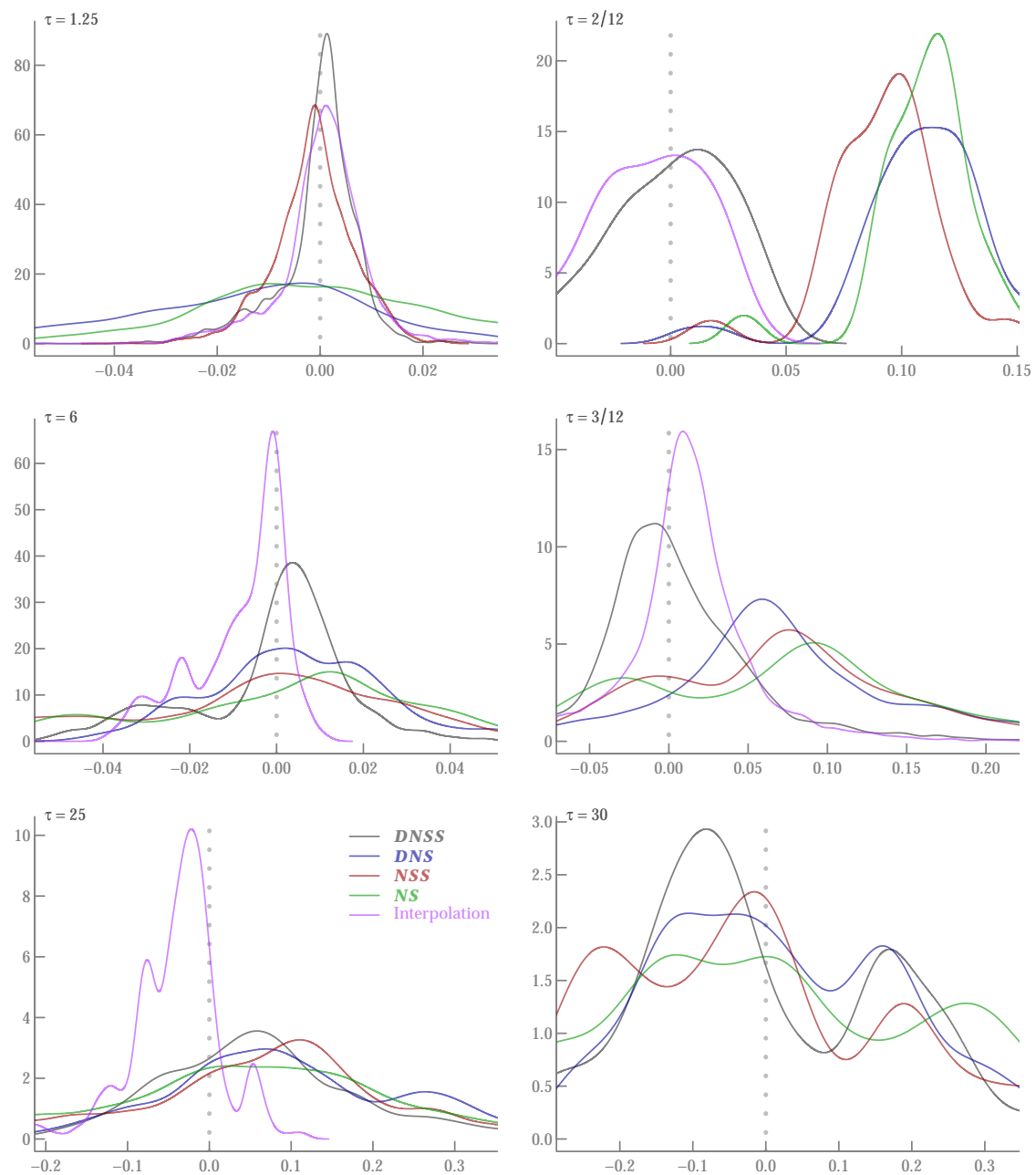


Figure A8. Kernel density estimates of prediction error distributions. Predictions from Bank of Canada yield data with maturities $\tau_{16} = (3/12, 6/12, 0.75, 1, 1.5, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30)$ (left panel) and U.S. Department of Treasury with maturities $\tau_9 = (1/12, 6/12, 1, 2, 3, 5, 7, 10, 20)$ (right panel). *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed. The daily sample is 4 January 2000 to 31 December 2018 (4719 observations) for Bank of Canada (left panel) and 31 July 2001 to 31 December 2018 (4357 observations some of which may be missing) for U.S. Department of Treasury (right panel).

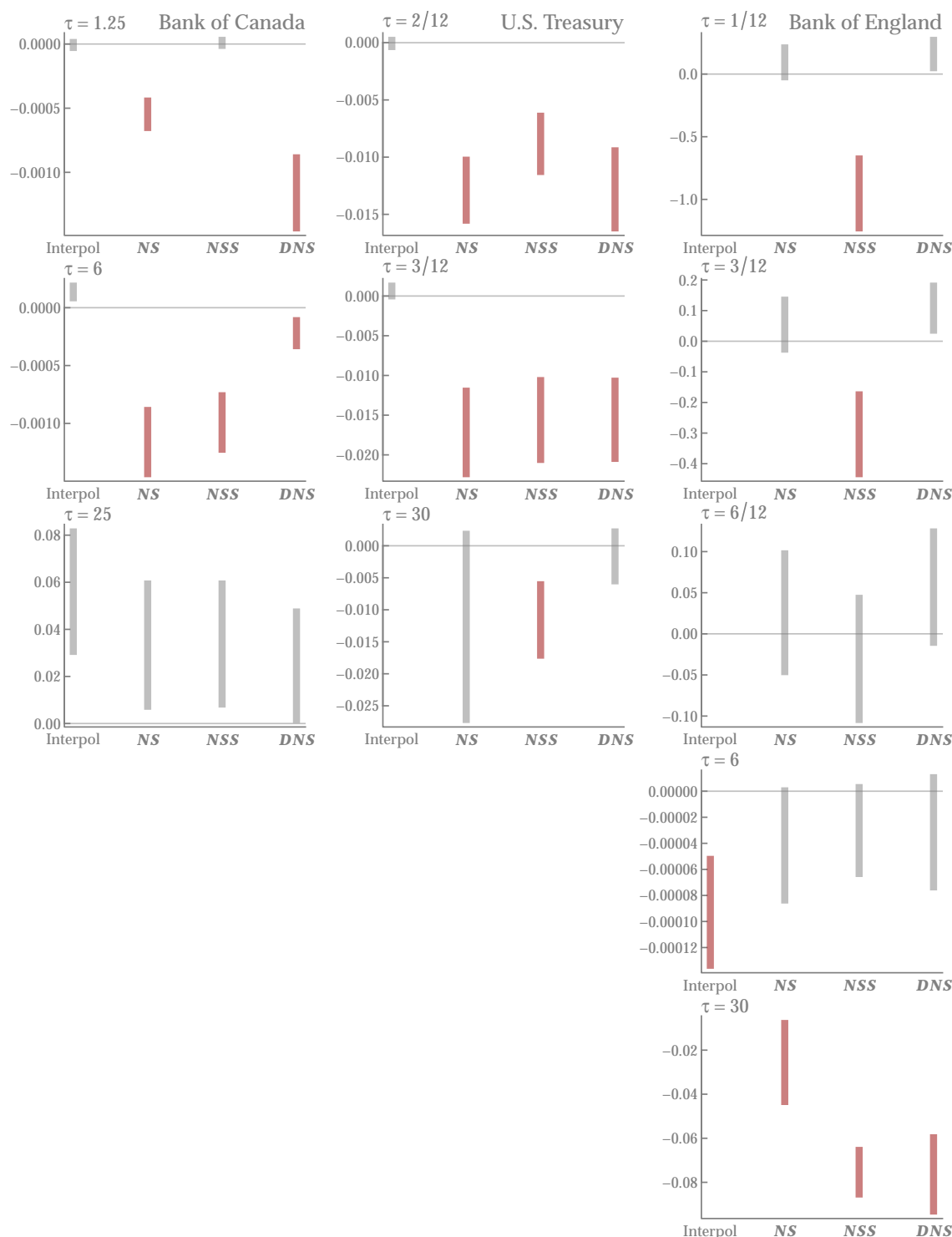


Figure A9. Prediction mean squared error (MSE) differences. Plotted are 0.95 confidence intervals for the difference in MSE between *DNSS* and other models. A negative difference indicates better forecasts for *DNSS* relative to the comparison model. Red intervals indicate statistical significance at size 0.05. *DNSS* is the dynamic Nelson-Siegel-Svensson specification where both decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ follow an AR(1) process, *DNS* is the dynamic Nelson-Siegel specification with a single decay parameter $\lambda_{1,t}$ that follows an AR(1) process, *NSS* is the Nelson-Siegel-Svensson specification with the two decay parameters λ_1 and λ_2 assumed to be fixed, *NS* is the Nelson-Siegel specification with a single decay parameter λ_1 assumed to be fixed.

References

1. Nyman-Andersen, P. *Yield Curve Modelling and a Conceptual Framework for Estimating Yield Curves: Evidence from the European Central Bank's Yield Curves*; Technical Report Statistics Paper Series 27; European Central Bank: Frankfurt, Germany, 2018.
2. Koo, B.; Vecchia, D.L.; Linton, O. Nonparametric Recovery of the Yield Curve Evolution from Cross-Section and Time Series Information. 2019, unpublished.
3. Nelson, C.R.; Siegel, A.F. Parsimonious Modeling of Yield Curves. *J. Bus.* **1987**, *60*, 473–489. [\[CrossRef\]](#)
4. Svensson, L.E. Estimating Forward Interest Rates with the Extended Nelson and Siegel Method. *Sver. Riksbank Q. Rev.* **1995**, *3*, 13–26.
5. Söderlind, P.; Svensson, L. New techniques to extract market expectations from financial instruments. *J. Monet. Econ.* **1997**, *40*, 383–429. [\[CrossRef\]](#)
6. Koopman, S.J.; Mallee, M.I.; der Wel, M.V. Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson–Siegel Model With Time-Varying Parameters. *J. Bus. Econ. Stat.* **2010**, *28*, 329–343. [\[CrossRef\]](#)
7. Pedersen, H.; Swanson, N.R. A survey of dynamic Nelson–Siegel models, diffusion indexes, and big data methods for predicting interest rates. *Quant. Financ. Econ.* **2019**, *3*, 22–45. [\[CrossRef\]](#)
8. Fisher, M.; Nychka, D.; Zervos, D. *Fitting the Term Structure of Interest Rates With Smoothing Splines*; Federal Reserve System: Washington, DC, USA, 1995.
9. Linton, O.; Mammen, E.; Nielsen, J.P.; Tanggaard, C. Yield curve estimation by kernel smoothing methods. *J. Econom.* **2001**, *105*, 185–223. [\[CrossRef\]](#)
10. Litterman, R.; Scheinkman, J. Common Factors Affecting Bond Returns. *J. Fixed Income* **1991**, *1*, 54–61. [\[CrossRef\]](#)
11. de Pooter, M. *Examining the Nelson–Siegel Class of Term Structure Models*; Tinbergen Institute Discussion Paper 07-043/4; Tinbergen Institute: Amsterdam, The Netherlands; Rotterdam, The Netherlands, 2007.
12. Filipovic, D. Exponential-polynomial families and the term structure of interest rates. *Bernoulli* **2000**, *6*, 1081–1107. [\[CrossRef\]](#)
13. Christensen, J.H.; Diebold, F.X.; Rudebusch, G.D. The affine arbitrage-free class of Nelson–Siegel term structure models. *J. Econom.* **2011**, *164*, 4–20. [\[CrossRef\]](#)
14. Duffie, D.; Kan, R. A Yield-Factor Model of Interest Rates. *Math. Financ.* **1996**, *6*, 379–406. [\[CrossRef\]](#)
15. Krippner, L. A Theoretical Foundation for the Nelson–Siegel Class of Yield Curve Models. *J. Appl. Econom.* **2015**, *30*, 97–118. [\[CrossRef\]](#)
16. Dai, Q.; Singleton, K.J. Expectation puzzles, time-varying risk premia, and affine models of the term structure. *J. Financ. Econ.* **2002**, *63*, 415–441. [\[CrossRef\]](#)
17. Christensen, J.H.; Diebold, F.X.; Rudebusch, G.D. An arbitrage-free generalized Nelson–Siegel term structure model. *Econom. J.* **2009**, *119*, C33–C64. [\[CrossRef\]](#)
18. Diebold, F.X.; Li, C. Forecasting the term structure of government bond yields. *J. Econom.* **2006**, *130*, 337–364. [\[CrossRef\]](#)
19. Swanson, N.R.; Xiong, W. Predicting Interest Rates Using Shrinkage Methods, Real-Time Diffusion Indexes, and Model Combinations. *J. Appl. Econom.* **2020**, *35*, 587–613. [\[CrossRef\]](#)
20. Coroneo, L.; Nyholm, K.; Vidova-Koleva, R. How arbitrage-free is the Nelson–Siegel model? *J. Empir. Financ.* **2011**, *18*, 393–407. [\[CrossRef\]](#)
21. Fearnhead, P.; Künsch, H.R. Particle Filters and Data Assimilation. *Annu. Rev. Stat. Appl.* **2018**, *5*, 421–449. [\[CrossRef\]](#)
22. Kutschireiter, A.; Surace, S.C.; Pfister, J.P. The Hitchhikers Guide to Nonlinear Filtering. *arXiv* **2019**, arXiv:1903.09247.
23. Julier, S.J.; Uhlmann, J.K. New extension of the Kalman filter to nonlinear systems. *Proc. SPIE* **1997**, *3068*, 182–193.
24. Künsch, H.R. Particle filters. *Bernoulli* **2013**, *19*, 1391–1403. [\[CrossRef\]](#)
25. Naesseth, C.A.; Lindsten, F.; Schön, T.B. Elements of Sequential Monte Carlo. *arXiv* **2019**, arXiv:1903.04797.
26. Schon, T.; Gustafsson, F.; Nordlund, P.J. Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Trans. Signal Process.* **2005**, *53*, 2279–2289. [\[CrossRef\]](#)
27. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [\[CrossRef\]](#)

28. Durbin, J.; Koopman, S. *Time Series Analysis by State Space Methods*, 2nd ed.; Oxford University Press: Oxford, UK, 2012.
29. Diebold, F.X.; Shin, M. Assessing point forecast accuracy by stochastic error distance. *Econom. Rev.* **2017**, *36*, 588–598. [[CrossRef](#)]
30. Jin, S.; Corradi, V.; Swanson, N.R. Robust Forecast Comparison. *Econom. Theory* **2017**, *33*, 1306–1351. [[CrossRef](#)]
31. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.
32. Andrieu, C.; Davy, M.; Doucet, A. Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions. *IEEE Trans. Signal Process.* **2003**, *51*, 1762–1770. [[CrossRef](#)]
33. Nemeth, C.; Fearnhead, P.; Mihaylova, L. Sequential Monte Carlo Methods for State and Parameter Estimation in Abruptly Changing Environments. *IEEE Trans. Signal Process.* **2014**, *62*, 1245–1255. [[CrossRef](#)]
34. Julier, S.J.; Uhlmann, J.K. Consistent debiased method for converting between polar and Cartesian coordinate systems. *Proc. SPIE* **1997**, *3086*, 110–121.
35. Julier, S.J.; Uhlmann, J.K.; Durrant-Whyte, H.F. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Autom. Control* **2000**, *45*, 477–482. [[CrossRef](#)]



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).