

# Bayesian Semiparametric Dynamic Nelson-Siegel Model \*

Cem Çakmaklı<sup>†</sup>

*Department of Quantitative Economics  
University of Amsterdam*

February, 2013

## Abstract

This paper proposes the Bayesian semiparametric dynamic Nelson-Siegel model, where the density of the yield curve factors and thereby the density of the yields are estimated along with other model parameters. This is accomplished by modeling the error distributions of the factors according to a Dirichlet process mixture. The semiparametric structure of the factors enables us to capture various forms of non-normalities including fat tails, skewness and non-linear dependence between factors using a unified approach. The potential of the proposed framework is examined using US bond yields data. The results show that the model can identify two different periods with distinct characteristics. While the relatively stable years of late 1980s and 1990s comprise the first period, the second period captures the years of severe recessions including the recessions of 1970s and 1980s and the recent recession of 2007-9 together with highly volatile periods of Federal Reserve's monetary policy experiments in the first half of 1980s. Interestingly, results point out a nonlinear dependence structure between the factors contrasting existing evidence.

**Keywords:** Dynamic factor model, Yield curve, Nelson-Siegel model, Dirichlet process mixture, Bayesian inference

**JEL Classification:** C14, C33, C38, G12

---

\*I thank to Richard Paap and Dick van Dijk for valuable comments and suggestions in various phases of this research. I thank to Peter Boswijk, Marco Del Negro, Domenico Giannone, Jim Griffin, Lennart Hoogerheide, Mark Jensen, Maria Kalli, John Maheu, Simon Potter and participants of the European Seminar on Bayesian Econometrics conference (ESOB 2011, Brussels, Belgium, November 2011) and seminar participants in University of Amsterdam for valuable comments and suggestions. Any remaining errors are my own.

<sup>†</sup>University of Amsterdam, Department of Quantitative Economics, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands, e-mail: C.Cakmakli@uva.nl.

# 1 Introduction

Modeling the evolution of large cross sections of treasury bond yields has been an intensive area of econometric research for many decades. Examining the patterns of bond yields is of key importance for many reasons including understanding the expectations of economic agents, managing fiscal debt, pricing bonds and related derivatives, pricing sovereign risk, and portfolio allocation. Typically, the common patterns observed in bond yields can be summarized using a few key factors. Not surprisingly, models that exploit this factor structure have been the major workhorse in modeling the common variation of treasury bond yields.

While the cross section of bond yields or put differently, the dependence structure of the yields on the underlying factors is well studied, the precise density of bond yields is not sufficiently explored yet. Typically, error distributions of the factors are assumed to be Gaussian in order to facilitate the econometric inference. However, relaxing the assumption of the Gaussianity and estimating the density explicitly is of crucial importance for various reasons. On the one hand, treasury bonds constitute an important class of financial assets. It is well known that the (joint) distribution of returns on many financial assets exhibits non-Gaussian behavior including fat-tails, excess skewness and nonlinear dependence structures. On the other hand, estimating the density of the multivariate bond yields explicitly is of key importance for uncovering the uncertainty around the point predictions and understanding the entire distribution of the bond yields with varying maturities.

In this paper, I propose a Bayesian semiparametric alternative to the assumption of Gaussianity for yield curve factors estimated using a popular class of yield curve models, the dynamic Nelson-Siegel model. More explicitly, I specify the distribution of the factors using a Dirichlet process mixture that is proved to be very useful in many applications of econometrics, see for example Chib and Hamilton (2002); H. (2002); Griffin and Steel (2004); Jensen (2004); Conley et al. (2008); Jensen and

Maheu (2010), among others. The resulting model does not rely on the assumption of Gaussianity of the factors nor I presume a specific distribution for the factors. Instead, using the semiparametric structure I estimate the factor density along with the factors. I propose an efficient and computationally tractable algorithm to obtain Bayesian inference. I analyze the potential of the Bayesian semiparametric dynamic Nelson-Siegel model using the panel time-series of US zero-coupon bond yields.

Traditional approaches of modeling the dynamics of the yield curve mostly focus either on the no-arbitrage relations in equilibrium to model the yield curve<sup>1</sup> or focus on functions that can approximate the observed yield curve well (see McCulloch, 1975; Nelson and Siegel, 1987, for example). For the latter, Nelson and Siegel (1987) (adjusted in Siegel and Nelson (1988)) provide a parsimonious three-component exponential approximation to the yield curve. This model has been heavily used in fitting and predicting the large cross sections of bond yields (see Svensson, 1994; Gurkaynak et al., 2007; Nyholm and Rebonato, 2008, for example).

However, the framework of Nelson and Siegel (1987) concerns only with fitting the yield curve of bonds for a given point in time. Diebold and Li (2006) extend this framework by imposing a first order autocorrelation structure on each component and thereby adding a time dimension. Diebold et al. (2006) show that the resulting dynamic model can be cast in a dynamic factor model (as a subclass of state-space models) framework with three factors, where these factors can be interpreted as the level, slope and curvature of the yield curve.<sup>2</sup> The ‘dynamic’ Nelson-Siegel model

---

<sup>1</sup>See Vasicek (1977); Cox et al. (1985); Hull and White (1990) for example. Duffie and Kan (1996) provide a multifactor affine (yields are affine functions of underlying factors) term structure model in this vein. Dai and Singleton (2000) classify the canonical representation of the affine no-arbitrage term-structure models and focus on three factor models. However, many issues plague the estimation of these models including the existence of local maxima each having a different economic interpretation, see Orphanides and Kim (2005), and overparametrization. Indeed, Dai and S. (2002) and Duffee (2002), among others, impose some further restrictions on model parameters to alleviate these problems.

<sup>2</sup>The interpretation of the yield curve factors as level, slope and curvature is due to the seminal paper of Litterman and Scheinkemann (1991). They use principle component analysis to extract the three common factors from a large set of US treasury bond yields and they interpret these factors as level, slope and curvature.

(DNS) can jointly capture both the cross sectional variation of the yields due to varying maturities and the dynamic evolution of the yield curve well, see the analysis in De Pooter (2007) among others. Consequently, the DNS model constitutes an extremely attractive class of yield curve models with desirable statistical properties, while still preserving a parsimonious model structure.

As the main focus in the DNS model is on modeling the dynamic evolution of the yields using the point predictions of Gaussian factors, the density of the factors is usually neglected. However, many studies document evidence of deviations from the Gaussian factor distributions due to several reasons. Interestingly, these studies focus mostly on a specific reason in isolation of potential other aspects. First, for US bond yields the volatility before the Great Moderation<sup>3</sup> had been much higher than the succeeding period. Indeed, Hautsch and Ou (2009) and Koopman et al. (2010) for the U.S. and Bianchi et al. (2009) for the U.K., among others, document the time variation in the volatility of the yields and/or factors. These studies mostly employ stochastic volatility (Jacquier et al., 2004; Omori et al., 2007) or GARCH (Bollerslev, 1986) type approaches to model fat tails and the time variation in the volatility, which may not be efficient for the typical monthly data of yields as these approaches are mostly suitable for higher frequency data. On top of this, for the US, beginning of 1980s are subject to aberrant observations and quite irregular dynamics of bond yields due to the monetary policy experiments of Federal Reserve (see Friedman, 1984, for details). Second, these approaches usually do not take nonlinear dependencies into account; hence, correlations are typically assumed to be constant. Junker et al. (2006), for example, use asymmetric copulas to model the nonlinear dependencies between the factors extracted from an arbitrage-free affine term structure model. Third, distributions of yields are likely to be skewed as nominal interest rates cannot be negative. On the one hand, evidence from

---

<sup>3</sup>Persistent decline in the volatility of many US macroeconomic and financial indicators after mid 1984 is referred as Great Moderation.

the recent global financial crises 2007-9 and the aftermath of it indicates that the ‘zero bound’ is likely to be a binding constraint for nominal yields as treasury bond yields in many countries fell essentially to their zero bound. From a statistical perspective, existence of the zero lower bound as a binding constraint leads to a skewed distribution for bond yields. On the other hand, extremely high bond yields observed in the beginning of 1980s (and never observed again) indicate that an upper bound for nominal yields does not exist and extreme shocks are more likely to occur in the right tail of the yields’ distribution.

Unfortunately, it is not a straightforward task to model the non-Gaussianity or to estimate the density of the yields due to the large cross section of the yields. When the non-Gaussianity is modeled for each bond yield separately, overparametrization is a potential problem that may complicate econometric inference.<sup>4</sup> One elegant and natural way to model this non-Gaussianity and to estimate the density of the series is exploiting the factor structure, i.e. estimating the density of the factors instead of the density of the individual yields. This facilitates econometric inference substantially by-passing the problem of the curse of dimensionality without losing much explanatory power.<sup>5</sup> This is especially the case for the DNS model where a few key factors explain the variation in the entire cross section of the yields quite successfully.

By estimating the distribution of the dynamic factors along with other model parameters the Bayesian semiparametric DNS model provides a unified approach that can take all features of bond yields into account. In the empirical application using US yield data, I show that the Bayesian semiparametric DNS model captures

---

<sup>4</sup>For some specific cases the model may still remain parsimonious. For example, when the true data generating process follows a multivariate  $t$  distribution, the model only requires a single additional parameter to estimate. Here I do not follow this approach as I explicitly focus on the density estimation without assuming a specific density for the yields.

<sup>5</sup>Indeed, using a similar intuition Patton and Oh (2011) model the nonlinear dependence in high dimensional models using factor copulas where factors are modeled using some non-Gaussian distributions.

two different periods of high and low volatility. While the period with low volatility coincides with the times of Great Moderation, the period with high volatility captures the times of severe recessions of 1970s together with the recent recession of 2007-9 and the period of Federal Reserve's monetary experiments in the beginning of 1980s. An interesting finding is that the model identifies the two historical peak points of the yields in the beginning of 1980s as outliers in a separate cluster leading to more precise estimates of the parameters. Moreover, estimates of the correlations reveal an interesting pattern in the sense that estimates using the models with the semiparametric structure generally reveal a higher degree of correlation compared to the benchmark with Gaussian innovations. However, during turbulent times correlation coefficients decrease substantially or switch signs indicating a nonlinear pattern. Finally, I perform several sensitivity analyses regarding to prior specifications and the model structure that confirm the robustness of the results.

The remainder of the paper is organized as follows. Section 2 describes the Bayesian semiparametric dynamic Nelson-Siegel model. I discuss the Bayesian estimation methodology together with the prior and posterior specifications in Section 3, with full details being provided in Appendix A. Section 4 displays and discusses the results of the model using US data.<sup>6</sup> Finally, Section 5 concludes.

## 2 Model specification

### 2.1 Dynamic Nelson-Siegel model

Fitting a parametric curve on the large set of yields that can approximate the observed yield curve successfully using only a few key parameters facilitates the analysis of the entire space of the yields. Following this idea, Nelson and Siegel (1987) pro-

---

<sup>6</sup>I examine the potential of the model and the algorithm using two simulated datasets. The results related to these simulated dataset are provided in the web appendix in my website <http://www1.fee.uva.nl/pp/personalpage.asp?personid=542&page=pd>.

vide a parsimonious yet flexible specification that can accommodate various forms of observed yield curves.

Explicitly, let  $y_t(\tau)$  be the yield of a zero-coupon bond at time  $t$  with the maturity date  $t + \tau$ . Nelson and Siegel (1987) (extended by Siegel and Nelson (1988)) specify the forward curve using an exponential approximation, which translates into the following specification of the yield curve with only three components<sup>7</sup> as

$$y_t(\tau) = f_{1,t} + f_{2,t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + f_{3,t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \quad (1)$$

Although (1) is governed by only four parameters;  $(\lambda, f_{1,t}, f_{2,t}, f_{3,t})'$ , the main advantage of this parsimonious specification is its flexibility to accommodate different forms of the yield curve successfully. The parameter  $\lambda$  in (1) governs the exponential decay rate determining the maximum loading of the third coefficient. Once the decay rate is determined, the key parameters  $(f_{1,t}, f_{2,t}, f_{3,t})'$  act as the main determinants of the cross section of yields.

The framework of Nelson and Siegel (1987), however, concerns only with fitting the yield curve of bonds for a given point in time. As a result, while in the classical specification of Nelson and Siegel (1987),  $(f_{1,t}, f_{2,t}, f_{3,t})'$  are interpreted as static coefficients to be estimated at a given point in time, Diebold and Li (2006) interpret these parameters as factors governing the evolution of the cross section of yields over time. The first factor loads on each yield equally and it can be interpreted as a level factor. Additionally, as  $\tau \rightarrow \infty$  the loadings of the second and third factors approach to zero leaving only the first factor to determine the long end. Hence, this factor can also be interpreted as long-term factor. The loading on the second factor starts with 1 at the short end. As  $\tau \rightarrow \infty$  it decays quickly to 0, thus, the effect of the second factor is mostly on the short end leading to the interpretation of this

---

<sup>7</sup>There is a one-to-one mapping between the forward curve and the yield curve as the yield is an equally weighted average of the forward rates.

factor as short-term factor. Furthermore, this factor is closely related to the yield curve slope as the change of it affects the short end more than the long end changing the slope of the yield curve. The loading on the third factor starts at 0 as  $\tau \rightarrow 0$  and increases as the maturity increases. After some maximum level it again decreases to 0 as  $\tau \rightarrow \infty$ . This implies that the third factor affects mostly the medium term maturities leading to the interpretation as a medium-term factor. Moreover, this effect in the medium term determines the curvature of the yield curve.

Given this interpretation of the Nelson-Siegel factors, the evolution of these factors shapes the evolution of the entire yield curve over time. Indeed, to explore this time variation Diebold and Li (2006) and Diebold et al. (2006) extend this framework by imposing a first order autocorrelation structure to these factors. The resulting state-space specification for time period  $t$  is given by

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} f_{1,t} \\ f_{2,t} \\ f_{3,t} \end{pmatrix} + \begin{pmatrix} \eta_t(\tau_1) \\ \eta_t(\tau_2) \\ \vdots \\ \eta_t(\tau_N) \end{pmatrix}, \quad (2)$$

and the evolution of the factors (state equation) takes the following form

$$\begin{pmatrix} f_{1,t} \\ f_{2,t} \\ f_{3,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \mu_{3,t} \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \\ f_{3,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}. \quad (3)$$

In the standard dynamic Nelson-Siegel model the error terms in both equations are distributed according to a multivariate Gaussian distribution. The resulting model



can be written in the compact notation as

$$\begin{aligned} y_t &= Hf_t + \eta_t & \eta_t &\sim N(0, R) \\ f_t &= Ff_{t-1} + \varepsilon_t & \varepsilon_t &\sim N(\theta_t) \text{ where } \theta_t = (\mu_t, Q_t) \end{aligned} \tag{4}$$

where  $\eta_t$  follows a multivariate normal distribution with mean zero and the covariance matrix  $R$ . The error distribution of the factors,  $\varepsilon_t$ , also follows a multivariate normal distribution (independent of  $\eta_t$ ) conditional on the model parameters  $\theta_t$ , where  $\theta_t$  includes the mean<sup>8</sup> and the variance of the factor error distribution in period  $t$ .

The key ingredient of the semiparametric model that differs from the conventional DNS model is the prior distribution used for the parameters  $\theta_t$ . In the standard case the use of a conjugate (Gaussian) or uninformative prior ensures the system to remain Gaussian. However, to estimate the density of the factors semiparametrically, I exploit Bayesian density estimation techniques, specifically, Dirichlet process mixtures by assuming an unknown prior distribution  $G$  for  $\theta_t$  where  $G$  itself is distributed according to a Dirichlet process.

## 2.2 Dirichlet process (prior) and Dirichlet process mixture

In this section, I introduce the semiparametric specification of the distribution of  $\theta_t$  using the Dirichlet process mixture, which leads to the estimation of the error distribution of factors. When the aim is estimation of a density using Bayesian inference, a prior distribution that spans a space of probability distribution functions that is large enough to capture all possible candidates is required. Consequently, Ferguson (1973) proposed Dirichlet process (prior) that is flexible enough to provide a prior specification satisfying this condition successfully.

Let  $G_0$  be a probability distribution function over some parameter space  $\Theta$  such

---

<sup>8</sup>The constant term in the state equation is suppressed in the error distribution as the mean of the error distribution.

that  $\theta \in \Theta$  and let  $\alpha$  be a positive scalar. A distribution  $G$  is distributed with Dirichlet process  $DP(G_0, \alpha)$  with base distribution  $G_0$  and precision parameter  $\alpha$  if for any measurable  $J$  partition of the parameter space,  $\theta_{1:J} = (\theta_1, \theta_2, \dots, \theta_J)$ , the distribution  $G = G(\theta_{1:J})$  follows a Dirichlet distribution  $Dir(\alpha G_0(\theta_1), \alpha G_0(\theta_2), \dots, \alpha G_0(\theta_J))$ . To clarify the underlying intuition and to demonstrate the functional form of the DP, I resort to the Polya's urn representation of the DP, see Blackwell and MacQueen (1973). Let us have an urn filled with a continuous sample of colored balls of size  $\alpha$  with the distribution of the colors of these balls denoted as  $G_0(\Lambda)$  that depends on the hyperparameters  $\Lambda$ . Suppose a ball is sampled from the urn and upon observing the color of the sampled ball another ball of exactly the same color is added the urn along with the sampled ball. Putting back the drawn ball in the urn keeps the initial sample size  $\alpha$  and the color distribution  $G_0(\Lambda)$  unchanged. However, as one more ball with identical color is added the urn the sample size increases by one. This implies that for the next round, either a ball with a new color from the initial sample of  $\alpha$  will be drawn (given that the base distribution  $G_0(\Lambda)$  is continuous) or the ball that is identical to the previously drawn ball will be drawn. Repeating this  $T$  times results in a  $T$  sample of balls (ignoring the initial sample size) with  $J$  different colors, where the distribution of this combination of  $J$  colors are distributed according to a Dirichlet distribution, see Ferguson (1973) for a formal proof.

When the state equation in the DNS model is considered, the parameters of the factor error distribution,  $\theta_t$ , are analogous to the balls of size  $T$  in the urn and the  $J$  partition of the parameters space,  $\theta_{1:J} = (\theta_1, \theta_2, \dots, \theta_J)$ , is analogous to the resulting  $J$  colors. The resulting Polya's urn scheme suggests that the conditional DP prior  $f(\theta_t | \theta_{-t}) | G = G \sim DP(G_0(\Lambda), \alpha)$  has the following form

$$f(\theta_t | \theta_{-t}) | G = G = \frac{\alpha}{\alpha + T - 1} G_0(\Lambda) + \sum_{i=1, i \neq t}^T \frac{\delta^j(\theta_i)}{\alpha + T - 1}, \quad (5)$$

where  $\delta(\cdot)$  is the Dirac function indicating the point mass such that  $\delta(\theta) = 1$  if  $\theta \in \theta_j$ . (5) indicates that at each step either a new  $\theta_t$  will be drawn from  $G_0(\Lambda)$  with probability proportional to  $\alpha$  or  $\theta_t$  will be a member of an existing cluster  $j$  with probability proportional to  $\sum_{i=1, i \neq t}^T \delta^j(\theta_i)$ .

As a flexible specification, Dirichlet process spans the entire distribution of the discrete probability functions with probability 1. This can be seen from the stick breaking representation of the DP, see Sethuraman (1994). The stick breaking representation corresponds to the following scheme. Suppose a part of  $W_j\%$  of a stick with a unit initial length is being constantly broken  $j$  times where  $W_j \sim \text{Beta}(1, \alpha)$ . Let denote the length of the broken parts as  $\pi_1 = W_1$ ,  $\pi_2 = W_2(1 - W_1)$  and  $\pi_j = W_j \prod_{i=1}^{j-1} (1 - W_i)(1 - W_2), \dots, (1 - W_{j-1})$ . As  $j \rightarrow \infty$  the unit length of the stick can be written as  $1 = \sum_{j=1}^{\infty} \pi_j$ . Since any partition can be the result of this stick breaking process, Dirichlet process spans the entire distribution of the discrete probability functions with probability 1. As the functional form also suggests, the length of the broken parts is analogous to the probabilities of each partition. Suppose each part of the stick, i.e. each partition, is denoted as  $\theta_j$ , which is a draw from the base distribution  $G_0$ . Accordingly, a draw of distribution  $G$  from the DP can be written as

$$G = \sum_{j=1}^{\infty} \pi_j \delta(\theta_j). \quad (6)$$

(6) points out the DP as an *infinite mixture* of point mass functions. Unfortunately, as the DP spans the entire distribution of the discrete probability functions with probability 1, the space of continuous probability functions is excluded. Although any distribution over  $\theta$  can be approximated arbitrarily accurately in the pointwise sense by a sequence of draws from the DP, the smoothness of the distribution is still of concern. Dirichlet process mixture model remedies this drawback by convolving the DP with a continuous kernel. This is accomplished by replacing the point mass function  $\delta(\theta_j)$  in (6) with a continuous function  $f(\cdot|\theta_j)$ , which is a continuous kernel

function conditional on the parameters  $\theta$  with the DP prior. This leads to a specification, denoted as the Dirichlet process mixture (DPM) that does span the entire set of continuous probability distributions with probability 1, see Antoniak (1974) for details. The DPM takes the form of

$$\sum_{j=1}^{\infty} \pi_j f(\cdot | \theta_j). \quad (7)$$

This implies that (7) corresponds to an (*countably*) *infinite number of mixtures* of underlying distributions and it provides a very intuitive and tractable way of estimation of unknown distributions.

Notice that, in practice the sample size is finite. This implies that the number of mixtures is limited with the sample size. Moreover, the discrete nature of the DP prior often results in a low number of mixture components that can successfully approximate the target density. This suggests the DPM as a very flexible yet parsimonious modeling approach (as the DNS model) to estimate an unknown density and to model the non-Gaussianities observed in the bond yields data.

### 2.3 Semiparametric Bayesian dynamic Nelson-Siegel model

When the distribution of the factors in (4) is of an unknown form, DPM structure provides a convenient and tractable way of estimation of this distribution. By assigning a DP prior for the parameters  $\theta_t$  in (4), the error distribution in state equation (factor evolution) is estimated along with other parameters by estimating the number of mixtures and mixture components. The resulting full model specification takes the following hierarchical form

$$\begin{aligned} y_t &= H f_t + \eta_t & \eta_t &\sim N(0, R) \\ f_t &= F f_{t-1} + \varepsilon_t & \varepsilon_t &\sim N(\theta_t) & \theta_t | G &\sim G & G &\sim DP(G_0(\Lambda), \alpha). \end{aligned} \quad (8)$$

The hierarchical representation indicates that the prior distribution  $G$  is distributed according to  $DP(G_0(\Lambda), \alpha)$  and together with the multivariate Gaussian density of the data conditional on  $\theta_t$  it forms the posterior distribution of  $\theta_t$  as a DPM. The standard DNS model with Gaussian innovations is a special case of the above specification with a single mixture component.

Without specifying the autoregressive coefficient matrix  $F$  and the covariance matrix of the measurement error  $R$  the model is not complete. One of the most prominent stylized facts of the yield curve models is the persistence in the factors. In many applications of the dynamic Nelson-Siegel model estimates of the autoregressive coefficient matrix  $F$  indicate uniformly this persistence with diagonal values above 0.9 or around it regardless of the model specification or estimation technique.<sup>9</sup> Following this evidence, I keep  $F$  out of the set of the parameters with DP prior.

Another parameter that I do not assign a DP prior is the covariance matrix of the measurement error,  $R$ . In fact, specifying a DPM structure in the measurement error distribution is a straightforward extension of my modeling approach. However, as the dimension of the yields is much higher than the dimension of the underlying factors such an approach would require many parameters to estimate and identification of the densities could be difficult. Alternatively, the semiparametric structure could be imposed only in the measurement equation rather than the state equation. My strategy, however, is to estimate the density of the yields and to model the observed non-Gaussianities in a parsimonious yet flexible specification using the underlying factors as discussed previously. Given that factors can explain the bulk of the variation in a large set of yields, I follow the approach of estimating the factor density.

---

<sup>9</sup>See for example Diebold and Li (2006); Diebold et al. (2006); Christensen et al. (2011); Koopman et al. (2010); Hautsch and Yang (2010), among others, for a comparison of the estimates of the autoregressive coefficient matrix.

### 3 Bayesian inference

In this section I provide details on econometric inference of the Bayesian semiparametric dynamic Nelson-Siegel model. The joint posterior distribution of the parameters with DPM distribution is constructed by the product of the DP prior together with the likelihood. In Section 3.1, I discuss the prior distributions. In Section 3.2, I discuss the full conditional posterior distributions related to the DPM and the resulting simulation scheme together with the DPM sampler. I discuss the posterior conditional distributions which do not follow a DPM distribution in Appendix A as these distributions are identical with the parameter distributions in the standard DNS model with Gaussian innovations. The Bayesian model selection procedures for model comparison are discussed in Section 3.4.

#### 3.1 Prior distributions

I start to the specifications of the prior distributions with the specifications related to the DP. This includes the specification of the base prior and the specification of the hyperprior for the precision parameter  $\alpha$ . For the base prior  $G_0$  I use a normal-inverse Wishart distribution as follows

$$G_0(\mu_0, a, \nu, S) = N - IW(\mu_0, a, \nu, S) \mapsto \mu|\Sigma \sim N\left(\mu_0, \frac{\Sigma}{a}\right) \text{ and } \Sigma \sim IW(\nu, S), \quad (9)$$

where  $E(\mu) = \mu_0$  and  $E(\Sigma^{-1}) = \nu S^{-1}$ . Moreover,  $S$  can be decomposed further into  $S = \nu\nu I_3$ , where  $I_n$  is the identity matrix of the dimension  $n$ . This implies that a priori I assume uncorrelated errors. With limited information on the correlation structure among the mixture parameters this is a natural assumption, see for example Conley et al. (2008) for a similar specification.

For the precision parameter of the DP,  $\alpha$ , I use the hyperprior distribution of the

form

$$f(\alpha) \propto \left(1 - \frac{\alpha - \alpha_{\min}}{\alpha_{\max} - \alpha_{\min}}\right)^p. \quad (10)$$

I take the power parameter  $p = 0.8$  and  $\alpha \in [0.1, 3]$ , see Conley et al. (2008) for details about this hyperprior. The power value implies that the prior has more probability mass on more parsimonious specifications. I prefer such a hyperprior specification since I want to keep a balance between the estimation of the density and model overfitting. As a high value of  $\alpha$  implies a high number of clusters, a more flexible specification could easily result in model overfitting.

The remaining model parameters do not involve a DP prior. For the prior distribution of the autoregressive parameters in (3), I assume the standard multivariate normal distribution as

$$f(\text{vec}(F)) = N(\mathbf{0}_9, I_9), \quad (11)$$

where  $\text{vec}(F)$  is the vec operator stacking each columns of  $F$  into a single column vector.  $\mathbf{0}_9$  is a vector of zeros of dimension nine. Unreported results of prior sensitivity analysis suggests that the likelihood information dominates prior information for  $F$  and thus, the prior specification is of limited importance for this parameter.

For the covariance matrix of the measurement error  $R$  in (2), I assume a noninformative diagonal structure as all cross sectional dependence is cast to the factor structure. For each individual elements a noninformative diagonal structure on the prior distribution of  $R$  implies

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} \quad \text{for all } n = 1, 2, \dots, N, \quad (12)$$

where  $\sigma_n^2$ 's are the diagonal elements of  $R$  and  $N$  is the number of the cross sections of yields, i.e number of the maturities.

### 3.2 Posterior simulation

The fact that the number of mixtures as well as the mixture components of the distribution of the factors are unknown leads to an intractable likelihood function. Instead, I set up an MCMC algorithm to sample from the full conditional posterior distributions. Specifically, I use Gibbs sampling together with data augmentation (see Geman and Geman, 1984; Tanner and Wong, 1987) to obtain posterior results. The resulting simulation scheme is as follows

1. Initialize the parameters by drawing  $\theta_t$  from the base prior  $G_0$  and  $f_t$  from standard normal distribution for  $t = 0, 1, \dots, T$ . At step ( $m$ ) of the iteration
2. Sample  $F^{(m)}$  from  $p(F|f_{0:T}^{(m-1)}, \theta_{1:T}^{(m-1)})$ .
3. Sample  $R^{(m)}$  from  $p(\sigma_n^2|y_{1:T}, f_{1:T}^{(m-1)})$  for  $n = 1, 2, \dots, N$ .
4. Sample  $\theta_{1:T}^{(m)}$  from  $p(\theta_t|f_0^{(m-1)}, \varepsilon_{1:T}^{(m-1)}, \theta_{-t}^{(m-1)}, F^{(m)})$  for  $t = 1, 2, \dots, T$ .
5. Sample  $f_0^{(m)}$  and  $\varepsilon_{1:T}^{(m)}$  from  $p(f_0, \varepsilon_{1:T}|y_{1:T}, \theta_{1:T}^{(m)}, F^{(m)})$ .
6. Sample  $\alpha^{(m)}$  from  $p(\alpha|\theta_{1:T}^{(m)})$ .
7. Repeat (2)-(6)  $M$  times.

Alternatively, steps (3) and (4) can be replaced by sampling  $\theta_t^{(m)}$  from  $p(\theta_t|y_{1:T}, \theta_{-t}^{(m-1)}, F^{(m)})$  which will be discussed in detail in the next section. Steps (2),(3), and (5) are common to many models in the Bayesian state-space framework, see for example Kim and Nelson (1999). Details of the full conditional distributions for these parameters are given in Appendix A. Here I focus on the DPM sampler in step (4) and on the posterior simulation of the precision parameter  $\alpha$  in step (6).



### 3.2.1 Sampling of error distribution parameters in the transition equation, $\theta_t$

Sampling the parameters with DPM distribution is not an easy task as both the factors and these parameters are unobserved, and  $\theta_{1:T}$  relates to the observed data through the unobserved factors. First, I start with the general idea of the DPM sampler as proposed in Escobar and West (1995) and then I show how this can be integrated into the state-space framework to obtain inference in the Bayesian semiparametric dynamic Nelson-Siegel (DNS-DPM) model.

The posterior distribution of the parameters with the DPM distribution,  $p(\theta_t|y_{1:T})$  does not have a closed form solution. This is due to the fact that obtaining the posterior distribution requires the integration of the all remaining parameters, denoted as,  $\theta_{-t}$ , out as  $p(\theta_t|y_{1:T}) = \int p(\theta_t, \theta_{-t}|y_{1:T})d\theta_{-t}$ . Unfortunately this is not possible except for some trivial cases or for very low dimensions. Instead, MCMC samplers can be exploited using the conditional distributions, see Escobar and West (1995). Notice that, the joint posterior distribution can be defined in terms of the conditionals as

$$p(\theta_{1:T}|y_{1:T}) = p(\theta_1|y_{1:T})p(\theta_2|\theta_1, y_{1:T}) \dots p(\theta_T|\theta_{1:T-1}, y_{1:T}). \quad (13)$$

Equation (13) implies that each of the parameters at time  $t$  can be sampled from a distribution conditional on the parameters sampled sofar as implied in (5) replacing  $T$  with  $t$ . Moreover, as the sequence of these parameters are exchangeable, each of the parameters can be treated as if it were the last observation leading to the posterior distribution of a single parameter as follows

$$p(\theta_t|\theta_{-t}, y_{1:T}) \propto p(y_{1:T}|\theta_t, \theta_{-t})p(\theta_t|\theta_{-t}). \quad (14)$$

Gibbs sampler (Geman and Geman, 1984) can be used to obtain a sample from

the joint posterior distribution (13) by replacing the conditionals with  $p(\theta_t|\theta_{-t}, y_{1:T})$  using (14). One iteration of the sampler to estimate  $\theta_t$  implies to draw from (14) conditional on the previous draws of  $\theta_{-t}$  and the data, see Escobar and West (1995). Unfortunately, (14) does not follow a standard distribution in this case due to the state-space structure preventing Bayesian inference using Gibbs sampling. A generic way to sample from the distribution in (14) is to setup a Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) with the conditional prior as the candidate distribution, see Neal (2000). This would require to evaluate the following acceptance probability  $\rho^{(m)}$  at iteration ( $m$ ) of the MCMC sampler

$$\rho^{(m)} = \min \left( 1, \frac{p(y_{1:T}|\theta_{1:T})p(\theta_t^*|\theta_{-t})q(\theta_t^*|\theta_{-t})}{p(y_{1:T}|\theta_{1:T})p(\theta_t^{(m-1)}|\theta_{-t})q(\theta_t^{(m-1)}|\theta_{-t})} \right). \quad (15)$$

Note that both  $p(\theta_t^*|\theta_{-t})$  and  $p(\theta_t^{(m-1)}|\theta_{-t})$  are identical as  $\frac{\alpha}{\alpha+T-1}G_0(\Lambda) + \sum_{i=1, i \neq t}^T \frac{\delta_{\theta_i}^j}{\alpha+T-1}$  and independent across draws. Hence, (15) corresponds to an independence chain MH algorithm with the following acceptance probability

$$\rho^{(m)} = \min \left( 1, \frac{p(y_{1:T}|\theta_t^*, \theta_{-t})}{p(y_{1:T}|\theta_t^{(m-1)}, \theta_{-t})} \right), \quad (16)$$

see Caron et al. (2008). The acceptance probability indicates that for sampling  $\theta_t$  a full Kalman filter step is required to evaluate the likelihood. This implies that drawing a single  $\theta_t$  requires  $O(T)$  operations. As  $\theta_t$  is sampled for  $t = 1, 2, \dots, T$ , a full sweep requires  $O(T^2)$  operations. To solve this problem Caron et al. (2008) suggest a decomposition based on Doucet and Andrieu (2001) to decrease the number of operations.<sup>10</sup>

A drawback of this procedure, however, is that the behavior of the MH algorithm may differ substantially from the Gibbs sampling algorithms with conjugate priors,

---

<sup>10</sup>This decomposition and the resulting algorithm is described in detail in the web appendix in my website <http://people.few.eur.nl/cakmakli/>.

see Neal (2000) for a discussion. One reason is that while other methods consider all mixture components when deciding for a new draw, for MH method it is more likely to consider changing the draw to a component with many observations rather than with few observations, because the candidate density is the DP prior. Changing a draw to a new component is proportional to  $\alpha$ ; hence, with small values of  $\alpha$  the probability of a new cluster may be inefficiently low. Notice that, even if a new cluster is sampled from the candidate density whether it will be accepted depends on the acceptance probability. This may dramatically slow down the mixing capability of the chain further unless a high value for  $\alpha$  is chosen, which is counter intuitive for many economic applications as this would imply tens (if not hundreds) of mixtures.

Closely related, the choice of the base prior plays a crucial role in the sense that the efficiency of the independence chain algorithm depends on the appropriate choice of the DP prior (candidate density) for the posterior distribution in (14). The reason is that convolving the DP prior with the data likelihood may result in a posterior that is quite different from the prior.

In this paper, I depart from this algorithm based on MH step and I follow a different path where direct sampling is possible. I decompose the likelihood function into the distribution of the yields conditional on the factor errors and the distribution of the factor errors conditional on the parameters with the DP prior. During the steps of the Gibbs sampler the state errors are estimated for each iteration of the sampler, and thus, for estimation of the  $\theta_t$ , these estimated state errors can be treated as observed data. That would facilitate the estimation of the  $\theta_t$  greatly as in this case posterior distribution has a closed form solution and MH algorithm is not required.

The resulting posterior distribution conditional on the factor errors becomes

$$\begin{aligned}
 p(\theta_t | \theta_{-t}, \varepsilon_{1:T}) &\propto p(\varepsilon | \theta_t) p(\theta_t | \theta_{-t}) \\
 &\propto \frac{\alpha q_{0,\varepsilon}}{\alpha + T - 1} p(\varepsilon | \theta_t) G_0(\theta_t) + \frac{\sum_{i=1, i \neq t}^T \delta_{\theta_i}^j}{\alpha + T - 1} p(\varepsilon | \theta_j),
 \end{aligned} \tag{17}$$

where  $q_{0,\varepsilon} = \int p(\varepsilon|\theta_t)G_0(\theta_t)d\theta_t$ ,  $p(\varepsilon|\theta_t)G_0(\theta_t)$  is the posterior distribution using the factor likelihood and the base prior, and  $p(\varepsilon|\theta_j)$  is the factor density conditional on the parameters in the  $j^{\text{th}}$  cluster. Fortunately, as I use factor errors instead of the observed yields the integral in  $q_{0,\varepsilon}$  has a closed-form solution. As  $\theta_t = (\mu_t, Q_t)$ ,  $q_{0,\varepsilon}$  takes the following form

$$\begin{aligned}
q_{0,\varepsilon} &= \int p(\varepsilon|\mu_t, Q_t)G_0(\mu_t, Q_t|\Lambda)d\mu_t dQ_t \\
&= \frac{p(\varepsilon_t|\mu_t, Q_t)p(\mu_t|Q_t)p(Q_t)}{p(\mu_t|\varepsilon_t, Q_t)p(Q_t|\varepsilon_t)} \\
&= 2^{\frac{3}{2}}(2\pi)^{-\frac{3}{2}}\left(\frac{a}{a+1}\right)^{\frac{3}{2}}\frac{\Gamma_3\left(\frac{\nu+1}{2}\right)}{\Gamma_3\left(\frac{\nu}{2}\right)}\frac{|V|^{\frac{\nu}{2}}}{\left|V+\frac{a}{a+1}(\varepsilon_t-\mu_0)(\varepsilon_t-\mu_0)'\right|^{\frac{\nu+1}{2}}}.
\end{aligned} \tag{18}$$

where  $\Gamma_3(\cdot)$  is the multivariate Gamma function of dimension 3 (the dimension of the factors), see, for example, Escobar and West (1995); Conley et al. (2008). To obtain a sample from the posterior distribution (17), let  $c_t$  be the indicator function taking the value  $j$  if observation in period  $t$  belongs to the  $j^{\text{th}}$  cluster. Following this, at step ( $m$ ) of the iteration, conditional on the factor errors,  $\varepsilon_{1:T}^{m-1}$ , and initial factors,  $f_0, \theta_t^m$  for  $t = 1, 2, \dots, T$  can be sampled using the following steps

1. For  $t = 1, 2, \dots, T$  sample  $c = (c_1, c_2, \dots, c_T)'$  from  $p(c_t|c_{-t}^{(m-1)}, f_0, \varepsilon_{1:T}^{(m-1)})$  for  $t = 1, 2, \dots, T$  using the Dirichlet process mixture posterior using (17),
2. For all  $c = c_j$  sample  $\theta_j$  from  $p(\theta_j|c^{(m)}, f_0, \varepsilon_{1:T}^{(m-1)}) \propto \prod_{i:c_i=j} [f(\varepsilon_i|\theta_j)] G_0(\theta_j)$ .

The second step of the algorithm is often called as the ‘remix’ step and it is to ensure that the values of  $\theta_j$  can change using all of the observations in that cluster rather than a single observation. This is accomplished by drawing it from an equivalent distribution to  $p(\theta_1, \dots, \theta_T|f_0, \varepsilon_{1:T})$  as in the second step, see West et al. (1994); MacEachern and Müller (1998).

### 3.3 Sampling DP precision parameter, $\alpha$

The posterior distribution of  $\alpha$ ,  $p(\alpha|J)$ , is the product of the hyperprior in (10) and the likelihood function  $p(J|\alpha)$  where  $J$  is the number of clusters. The resulting posterior is

$$p(\alpha|J) \propto p(\alpha)T!\alpha^J \frac{\Gamma(\alpha)}{\Gamma(\alpha + T)}, \quad (19)$$

see Antoniak (1974). Given (19) the gridgy Gibbs sampler approach of Ritter and Tanner (1992) can easily be implemented, see also Conley et al. (2008). Given that  $\alpha \in [0.1, 3]$  a grid in this interval can be setup based on the desired precision about the value of  $\alpha$ .

Escobar (1994) shows that the expected number of clusters for a given  $\alpha$ ,  $E(J|\alpha) = \sum_{i=1}^T \frac{\alpha}{\alpha+i-1}$  approaches to  $\alpha \ln\left(\frac{\alpha+T}{\alpha}\right)$  as  $T \rightarrow \infty$ . This implies that the interval for  $\alpha$  corresponds to the number of clusters to be between 1 and 16 with more support for lower values.

### 3.4 Model selection

The Bayesian semiparametric dynamic Nelson-Siegel model encompasses the standard DNS model with Gaussian innovations as a specific case. Hence, comparison of the standard model as a restricted version of the semiparametric model with the unrestricted semiparametric model would indicate whether there is a need for the semiparametric structure. Following this aim, in this section I discuss several Bayesian model comparison techniques.

Since the key task in my methodology is the density estimation I use the predictive likelihood as a first metric for model comparison. The one-step ahead predictive likelihood of the observation at  $t_0 + 1$ ,  $y_{t_0+1}$ , conditional on the previous observations

$y_{1:t_0}$ , is given by

$$f(y_{t_0+1}|y_{1:t_0}) = \int p(y_{t_0+1}|f_{t_0+1}, \theta, \alpha) p(f_{t_0+1}, \theta, \alpha|y_{1:t_0}) df_{t_0+1} d\theta d\alpha. \quad (20)$$

This can be computed as by first generating an ergodic sequence of  $\{f_{t_0+1}\}_{m=1}^M$ . For generating this sequence, first  $\theta_{t_0+1}^m$  is drawn from  $f(\theta_{t_0+1}^m|\theta_{1:t_0}^m, \alpha^m) = \frac{\alpha^m}{\alpha^m + t_0 - 1} G_0(\Lambda) + \sum_{i=1}^{t_0} \frac{\delta^j(\theta_i)}{\alpha^m + t_0 - 1}$  and using this draw  $f_{t_0+1}^m$  can be generated using (3). Next, the predictive likelihood of the observation at  $t_0 + 1$ ,  $PL(t_0 + 1)$ , can be computed as

$$f(y_{t_0+1}|y_{1:t_0}) \approx \frac{1}{M} \sum_{m=1}^M p(y_{t_0+1}|f_{t_0+1}^m, \theta_{1:t_0}^m, \alpha^m), \quad (21)$$

where  $p(y_t|f_{t_0+1}^m, \theta_{1:t_0}^m, \alpha^m)$  is a multivariate normal density implied in (8) and  $M$  is a sufficiently large number. A feature of the predictive likelihoods is that these can be used to compute the marginal likelihood as

$$p(y_{1:t_0}) = \prod_{t=1}^{T} f(y_{t+1}|y_{1:t}). \quad (22)$$

This provides a tool to analyze the contribution of each observation at time period  $t$  to the (log) marginal likelihoods as in (22), see Geweke and Amisano (2010).

As a second metric for evaluation of the models, I use the posterior predictive p-value ( $ppp$ ). The  $ppp$  provides a framework to examine whether the data simulated using the estimated model can replicate various features of the actual data. These features of the data can usually be formulated as functions of data,  $g(y)$ , such as the skewness or the kurtosis measures. If this is the case, it implies that indeed the model specification can approximate the underlying data generating process of the actual data well. As my aim is to capture the non-Gaussian features of the bond yields data, this metric is very suitable for comparison of the competing models.

The posterior predictive p-value can be computed as

$$\begin{aligned} p(g(y^{sim})|y_t) &= \int p(g(y^{sim})|y_t, \Gamma)p(\Gamma|y_t) \\ &= \int p(g(y^{sim})|\Gamma)p(\Gamma|y_t), \end{aligned} \tag{23}$$

where  $\Gamma$  denotes the parameter space. Computing (23) requires to generate simulated data,  $y^{sim}$ , using the posterior distribution of model parameters. Specifically, I use the distributions

$$\begin{aligned} f_t &\sim N(\mu_t + F f_{t-1}, Q_t) \\ y_t &\sim N(H f_t, R), \end{aligned} \tag{24}$$

successively to generate simulated data. The second equation in (23) stems from the fact that conditional on the model parameters, observations do not provide additional information for the simulated data. The *ppp* can be computed as

$$\begin{aligned} ppp &= p(g(y_t^{sim}) \geq g(y_t)) \\ &\approx \frac{1}{S} \sum_i^S \mathbf{1}[g(y_t^{sim}) \geq g(y_t)]. \end{aligned} \tag{25}$$

A *ppp* value close to 0.5 indicates that the model can generate the features of the actual data well, while a *ppp* value close to 0 or 1 implies the opposite.

## 4 Empirical results using US yield data

In this section I examine the potential of the semiparametric framework using a real bond yields dataset. Following this aim, in my empirical investigation I use the US treasury bond data and accordingly, I investigate the US yield curve. This dataset is analyzed extensively in many applications of the dynamic Nelson-Siegel model, and thus, it provides ample opportunities to compare my results with the standard applications of dynamic Nelson-Siegel model. The data consists of all Treasury bills, noncallable notes and bonds over the period January 1970 through December 2009

from the CRSP Monthly Treasury Cross-Sectional File.<sup>11</sup> The resulting dataset consists of 22 yields with maturities 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 60, 72, 84, 96, 108, 120 months and each of the yields has 480 monthly observations, i.e.  $N = 22$  and  $T = 480$ .

While there are some evidence on the variation in the volatility of the US yields, see for example Hautsch and Ou (2009); Koopman et al. (2010), the evidence for the changing intercept is limited. This leads me to consider a specific case of DNS-DPM model, where only the variance parameter is assumed to follow a DPM but the intercept parameters are assumed to be constant. This also enables me to evaluate the marginal contributions of modeling intercept and volatility semiparametrically. Following Jensen and Maheu (2010), I denote this model as DNS-DPMP. The estimation of this model is identical to the DNS-DPM model explained in detail in Section 3 with only (18) replaced by

$$q_{0,f} = 2^{\frac{3}{2}}(2\pi)^{-\frac{3}{2}} \frac{\Gamma_3\left(\frac{\nu+1}{2}\right)}{\Gamma_3\left(\frac{\nu}{2}\right)} \frac{|V|^{\frac{\nu}{2}}}{|V+(\varepsilon_t-\mu)(\varepsilon_t-\mu)'|^{\frac{\nu+1}{2}}}, \quad (26)$$

and the state equation in (3) replaced by

$$\begin{pmatrix} f_{1,t} \\ f_{2,t} \\ f_{3,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \\ f_{3,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}. \quad (27)$$

To limit the computational burden and to keep the model more parsimonious, I restrict the off-diagonal elements to be zero in the main analysis using the DNS framework, see Hautsch and Yang (2010) for similar restrictions.

Specifying the decay parameter  $\lambda$  is necessary to complete the DNS model spec-

---

<sup>11</sup>The instruments with special liquidity problems, which are the bills with less than one month to maturity and all notes and bonds with less than one year to maturity are eliminated, see Diebold and Li (2006) for details. Forward rates are computed using the unsmoothed Fama-Bliss method (see Fama and Bliss (1987)). These forward rates are converted into unsmoothed Fama-Bliss zero yields. I thank Michel van der Wel for making the data available to me.



ification. In many applications of the Nelson-Siegel model the value of  $\lambda$  is fixed to 0.0609 or 0.077, see for example Diebold and Li (2006); Diebold et al. (2006); Yu and Zivot (2011). The value 0.0609 (0.077) maximizes the loading on the curvature, at exactly 30 months (23.3 months). Fixing the value of  $\lambda$  prior to estimation facilitates the estimation of the model without losing much explanatory power. As the focus of the paper is on estimating the density of the factors I follow this common practice by fixing  $\lambda$  at 0.0609. Indeed, the sensitivity analysis provided at the end of this section, where I estimate  $\lambda$  along with other parameters reveals similar conclusions.

The estimates of the factors using the DNS-DPM model and the differences of these estimates from those of the DNS-DPMP and DNS-N models are displayed in Figure 1. The first graph in the figure shows the factor estimates using the DNS-DPM model. The patterns of the estimated factors are in line with previous findings in the sense that the level factor closely tracks the US inflation patterns and the slope factor exhibits some cyclical variation closely related to the business cycle, see the evidence in Diebold et al. (2006) for example. The differences of the estimates from the competing models reveal an interesting pattern. While the differences between the factor estimates of the DNS-DPM model and the DNS-DPMP model are limited the same does not hold for the DNS-N model specifically during the economic downturns.

[Insert Figure 1 about here]

In Figure 2, I display the course of the posterior standard deviations of the factors through the sample period. Interestingly, there is a unique pattern for all three factors in the sense that posterior standard deviations of the factors are much lower for the models that use the DPM structure compared to the standard model. Moreover, for the level and curvature factor, the precision is even higher for the DNS-DPM model compared to the DNS-DPMP model. This shows the increasing

precision of the factor estimates when the distribution of the factors is modeled semiparametrically using the DPM.

[Insert Figure 2 about here]

The parameter estimates of the competing models are shown in Table 1 with the variance of the measurement errors displayed in Table 2. The evolution of the parameters that follow the DPM and comparison of those with the estimates of the competing models are displayed in Figure 3.

[Insert Table 1 about here]

[Insert Table 2 about here]

[Insert Figure 3 about here]

First, the estimates of the autoregressive parameters are increasing when the DPM structure is used. Despite the increasing persistence, or put differently, despite the higher values of the estimates of the autoregressive parameters, the posterior standard deviations of these parameters are much lower than those in the standard model. This leads to the conclusion that the autoregressive parameters can be estimated much precisely using the model with DPM structure.

Second, the same also applies to the estimates of the intercept parameters when these are considered for the DNS-N and DNS-DPMP model in the sense that posterior standard deviations of the intercept parameters are much lower when DNS-DPMP model is considered. The increase in the persistence of the level and slope factors is accommodated with lower values (in absolute terms) in the intercept estimates leading to a similar mean level for these factors when these two models are considered. For the curvature factor, although the persistence increases, the intercept also increases in absolute terms leading to a mean that is higher than (in absolute terms) its counterpart using DNS-N model.

Third, the estimates of the variances of the measurement errors in (2), i.e. the diagonal elements of the  $R$  in Table 2 indicates an increasing precision when the semi-parametric structure is used. When the DNS-N model is compared to DNS-DPMP and DNS-DPM models it is seen that the variances is decreasing significantly in the short end of the yield curve when the semiparametric structure is used. While for the medium term yields variances are almost identical, there is only a marginal increase in variances in the long end with the use of the semiparametric structure. This implies that yield curve factors can explain even a greater portion of the variation of the entire space of the yields when the semiparametric structure is used.

When the parameters with DPM distribution are analyzed in Figure 3, it is seen that the estimates of the parameters related to the covariance matrix of the factor errors are very similar for the DNS-DPM and DNS-DPMP model. Interestingly, estimates of the intercept parameters for the DNS-DPM model are fluctuating around the value estimated by the DNS-DPMP model, while the estimate of the DNS-N model is greater than that of the DNS-DPMP accommodating these fluctuations. Nevertheless, these fluctuations are limited in magnitude with the only exceptions around 1980 which are the two extreme peaks of the historical interest rates observed during the change of the Federal Reserve's monetary target. It seems that the DPM model categorizes these observations as a separate cluster leading to the interpretation of these observations as outliers. Volatility estimates are in line with the existing evidence in the sense that the effect of the Great Moderation and the change in the Federal Reserve's monetary policy together with the effects of the recessions of 1970s, 1980s, and the recent 2007-9 crisis are nicely captured by the changing volatility with both DNS-DPM and DNS-DPMP models. The volatility of the curvature factor indicated by the DNS-DPMP model is slightly higher than the volatility indicated by the DNS-DPM model. This is due to the fact that part of the volatility is captured by the variation in the intercept parameter. Similar to the

estimates of the intercept the estimate of the volatility around 1980 is extremely high (the estimates in the figure is truncated to increase the visual quality of the graph) again leading to the interpretation of these observations as outliers. As a result of this changing volatility, the volatility estimate of the DNS-N model accommodates this with a volatility that is higher than the estimates of the volatility of the DPM models during the periods of Great Moderation and with a lower volatility during volatile periods.

The estimates of the correlation coefficients show that while the estimates of the DNS-N model are close to zero, the estimates of the DNS-DPM (DNS-DPMP) model are around 0.30 (0.30) for the correlation between the level and slope factors' error terms and around 0.20 (0.10) for the correlation between the level and curvature factors' error terms. An interesting finding is that these correlations decrease significantly or even changing signs during the volatile periods. This implies that during turbulent times the correlation between the long term yields (level factor) and the short and medium term yields (slope and curvature factors) are decreasing (or changing signs for the level-curvature correlation). As the turbulent times mostly correspond to the recessions, the dynamics of the yields with different maturities changes as the short term yields and corresponding spreads are more affected by the changes in the monetary policy to stimulate the economy while longer yields are affected mostly by the inflation expectations in the longer term, see Diebold et al. (2006), Ang et al. (2008) and Christensen et al. (2010) among others. The correlation coefficients of the slope and curvature factors' error terms are closer to zero with limited fluctuations. Again the extreme observations of 1980s enter here as a negative correlation during these periods unlike the other periods. These results show a nonlinear dependence structure between the factors, which is nicely captured using the semiparametric structure. As in the volatility case, the estimate of the DNS-N model balances between the changing correlations leading to moderate estimates of

the correlations.

I display the estimates of the precision parameter  $\alpha$  and the average number of clusters together with the marginal likelihood values for for 3, 12, 24, 60 and 120 months maturity yields in the bottom panel of Table 1. The estimates of  $\alpha$  for the DNS-DPMP and DNS-DPM models are very close to each other indicating that indeed the second moment is the main driving force of the DPM structure. In both cases the estimate of  $\alpha$  is close to one resulting in the average number of clusters around six. Unreported results show that the mode of the distribution of the number of clusters with more than 60 observations (5 years in total) is two indicating that on average there are two major periods with different characteristics and most of the remaining clusters capture aberrant observations. Interestingly, in many draws of the MCMC sampler the recent 2007-9 crisis is identified as a separate cluster different than the period before mid 1980s indicating a new period (cluster) though more information is required for more conclusive results.

Marginal likelihood values for the yields with 3 months maturity point a dramatic increase when switching from the DNS-N model to the DNS-DPMP model and a limited increase when switching from the DNS-DPMP model to the DNS-DPM model. When medium term maturities are considered similar dramatic increases in the marginal likelihood values can be observed when switching from the DNS-N model to the DNS-DPMP model. A striking result from the graphs as well as from the final marginal likelihood values is about the fairly large differences between the marginal likelihood values of the DNS-DPMP model and the DNS-DPM model for these maturities. It seems that modeling changes in the level of the factors is very important most notably for the curvature factor as this factor mostly affects medium term yields. When the longest maturity in the sample, i.e. yields with 120 months maturity, is analyzed it is seen that the increase in the marginal likelihood values is still high when switching from the DNS-DPMP model to the DNS-DPM model,

albeit not as large as for the medium maturities. This indicates the necessity of a flexible modeling approach that can capture various different characteristics of the US sovereign bond yields.

[Insert Figure 4 about here]

In Figure 4 I display the evolution of the (log) marginal likelihood values of the models for 3, 12, 24, 60 and 120 months maturities by computing the cumulative sum of the one-period ahead log predictive likelihood as stated in (22) over an expanding window starting from May 1980. The reason for this specific choice of the starting period is that the outliers prior to this period cause the predictive likelihood for the DNS-N model to be substantially low preventing a comparison of the models (as the marginal likelihood for the DNS-N model is then extremely low compared to other models) at the beginning of the sample. All five graphs has a common feature. The differences between the marginal likelihoods of the semiparametric models and DNS-N model increase always during the turbulent times of FED's monetary policy experiments in the beginning of 1980s as well as during the recent economic downturn called as 'Great Recession'. This clearly shows that the semiparametric models can capture the erratic behavior in the data successfully. This comparison between the models shows that when some parts of yields data exhibit non-conventional changes it may cause to ill-behaved regions in the posterior distributions blurring the econometric inference. Semiparametric structure in the DNS framework nicely captures this part as outliers or as a separate cluster with different dynamics, while the inference about the other parts remain unaffected. This tendency can be observed clearly in the graph for 3 months maturity yields. While the changes in the marginal likelihoods are substantial for turbulent times, it remains marginal for the times surrounding the Great Moderation.

As a final measure of model fit I display the posterior predictive p (*ppp*) value of the models for each maturity in Table 3. The objective functions are the sample

kurtosis and skewness as I want to measure how successfully the model can capture the non-Gaussian characteristics of the data.

[Insert Table 3 about here]

The results for the DNS-N model show that while the state-space structure can still capture the kurtosis of the data for various maturities to some extent it cannot do so for the skewness of the data as many of the  $ppp$  values are close to 0. On the contrary, semiparametric models perform quite well in capturing both the kurtosis and skewness. Additionally, the DNS-DPM model can capture the skewness of the yields better than the DNS-DPMP model although the increase in  $ppp$  values are limited. This is due to the fact the DNS-DPM specification models the intercepts semiparametrically allowing for skewed distributions.

#### 4.1 Sensitivity check of DP hyperparameters and the decay parameter

The estimation of the Bayesian semiparametric dynamic Nelson-Siegel model so far involves the estimation of the precision parameter  $\alpha$  but does not involve the estimation of the hyperparameters,  $\Lambda$ , of the base prior  $G_0$  and the estimation of the decay parameter  $\lambda$ . Estimation of the hyperparameters is a straightforward extension of the model given that conjugate hyperpriors on these parameters can easily be assigned. However, as the number of observations in this likelihood function related to the hyperparameters is the number of clusters, the informative content of the likelihood is limited and a careful selection of the hyperprior distribution is crucial. Instead, I fix the hyperparameters and perform a prior sensitivity analysis, see Jensen and Maheu (2010) for a similar analysis. I, therefore, re-estimate the model using the following six different hyperparameter settings

- Prior 2:  $G_0(\mu_0, a, \nu, S) = N - IW(\mu_0, 10, 5, 5 * 1 * I_3)$

- Prior 3:  $G_0(\mu_0, a, \nu, S) = N - IW(\mu_0, 15, 5, 5 * 1 * I_3)$
- Prior 4:  $G_0(\mu_0, a, \nu, S) = N - IW(\mu_0, 5, 10, 10 * 1 * I_3)$
- Prior 5:  $G_0(\mu_0, a, \nu, S) = N - IW(\mu_0, 5, 15, 15 * 1 * I_3)$
- Prior 6:  $G_0(\mu_0, a, \nu, S) = N - IW(\mu_0, 10, 10, 10 * 1 * I_3)$
- Prior 7:  $G_0(\mu_0, a, \nu, S) = N - IW(\mu_0, 15, 15, 15 * 1 * I_3)$ .

Posterior results are displayed in Table 4. Results show that while autoregressive parameters are in general not so sensitive to the different settings of hyperparameters, the parameter of the curvature is marginally sensitive to the choice of the degrees of freedom  $\nu$  of the variance parameter. With the increasing  $\nu$  the autoregressive parameter of the curvature is decreasing, albeit only marginally. It seems that the choice of  $a$  does not influence the results as changes in the value of this parameter are not as effective as changes in the value of  $\nu$ . A similar pattern can also be observed for the estimates of the precision parameter  $\alpha$  and the number of clusters. While changes in  $a$  do only have a minor effect on these estimates, the effect of the changes in  $\nu$  is relatively higher. Notice that, an increase in  $\nu$  implies a more informative prior on the variance structure causing to less variation in the clusters, i.e. to fewer number of clusters. Still, this effect is limited in the sense that in most extreme cases the mode of the cluster number distribution decreases only from six to four. As a result the estimates of  $\alpha$  also decrease to a limited extent. Overall, results are quite robust to different parameter settings.

In the estimation of the DNS models I fixed the decay parameter  $\lambda$  in order to facilitate the density estimation of the factors. Although the smooth factor loading structure of the DNS model provides superior identification for yield curve factors, still, elaboration of this may provide further insights especially on the effects for factor density estimation. Therefore, as a last sensitivity check I estimate a DNS-DPM model together with the  $\lambda$ , which is denoted as DNS-DPM- $\lambda$  model. Details



on the estimation of  $\lambda$  is provided in Appendix A. The results related to this model are displayed in Table 5 while the results related to the measurement errors and  $ppp$  values are displayed in the last column of the Table 2 and Table 3 respectively.

[Insert Table 5 about here]

The  $\lambda$  parameter is estimated as 0.069 with a posterior standard deviation of 0.011, thus the frequently used values 0.0609 and 0.077 lie in the 95% highest posterior density interval. Results in Table 5 indicates that the autoregressive parameters are quite similar to the DNS-DPM. The estimate of  $\alpha$  and accordingly the estimated number of clusters (posterior means) increase marginally, while the posterior modes are identical when the number of clusters are considered. Unreported results show that the dynamics of the parameters with DPM structure are identical to that of the DNS-DPM model. Hence, estimation of the  $\lambda$  hardly effects the DPM structure of the model. The results about the variance of the measurement errors in the last column of the Table 2 indicate that while there is a minor deduction in the variance when the first three maturities are considered at the short end, on the contrary, an increase in the variance is observed at the long end. One reason for this observation is the increasing uncertainty with the estimation of the  $\lambda$ . Another reason especially for the difference in the short and long end is due to the fact that the estimate of the  $\lambda$  is slightly larger than the fixed value. While small values of the decay parameter produce slow decay and provide a better fit to the curve at the long end, large values produce fast decay and provide a better fit to the curve at the short end. This effect can also be seen in the last column of the Table 3, where the  $ppp$  values are displayed. While the  $ppp$  value for the skewness indicates a better fit to the actual data for the shortest end this reverses for the long end as a result of a greater  $\lambda$  estimate than the fixed value. Overall, the changes are minor when  $\lambda$  is also estimated along with other model parameters and results related to the DPM structure remain unchanged.

## 5 Conclusion

In this paper, I put forward the Bayesian semiparametric dynamic Nelson-Siegel model to estimate the density of the yield curve factors in the dynamic Nelson-Siegel model. Estimating the density of the factors explicitly enables to model various characteristics of the US sovereign bond yields such as skewness, fat tails and nonlinear dependence between the underlying factors using a unified approach. Specifically, I model the distribution of the underlying factors semiparametrically using Dirichlet process mixtures. I provide an algorithm that is easy to implement and that makes use of the full conditional posterior distribution. Consequently, the sensitivity of the algorithm to the choice of the prior distributions is very limited.

In the empirical application I show that this unified methodology of the Bayesian semiparametric dynamic Nelson-Siegel model captures several aspects of the US treasury bond yields data. I locate two different periods with different characteristics of volatility dynamics. Moreover, the results indicate that the correlation between the errors of the factors are usually higher than the correlations indicated by a standard dynamic Nelson-Siegel model. However, this dependence decreases or switches signs during turbulent periods pointing out a nonlinear dependence structure between the factors.

The model and the resulting algorithm proposed in this paper can be used for other applications as well. In fact, the model falls into the class of the dynamic factor models, which in turn is a special case of the general state-space models. As the Bayesian inference as well as the accompanying algorithm are not dependent on the specific Nelson-Siegel factor loadings structure, several applications of the state-space models can be extended to allow for semiparametric structure in the state equation. Therefore, the methodology in this paper provides a unified approach to model different types of non-Gaussian behavior in a wide variety of applications of state-space models.

# Appendix A Full conditional posterior distributions

In this section I describe the full conditional posterior distributions of the parameters that are common to both the model with DPM structure and the model with Gaussian innovations. Conditional on the remaining parameters these distributions can be obtained by the product of the prior distributions and the likelihood function described in Section 3.1 and 3.2.

## A.1 Sampling of autoregressive parameters, $F$

Conditional on the states  $f_{0:T}$  and the parameters  $\theta_{1:T}$  the state equation in (3) can be rewritten as a Vector AutoRegression (VAR) of first order as follows

$$f_t = \mu_t + Ff_{t-1} + \varepsilon_t \text{ and } \varepsilon_t \sim N(0, Q_t). \quad (\text{A.1})$$

Redefining the variables such that

$$\begin{aligned} z_t &= f_t - \mu_t & \text{for } t = 1, \dots, T \\ x_t &= f_{t-1} & \text{for } t = 0, \dots, T-1, \end{aligned} \quad (\text{A.2})$$

(A.1) can be rewritten as

$$z_t = Fx_t + \varepsilon_t \text{ and } \varepsilon_t \sim N(0, Q_t). \quad (\text{A.3})$$

In compact form the model becomes

$$Z = FX + U, \quad (\text{A.4})$$

where  $Z = (z_1, z_2, \dots, z_T)$ ,  $X = (x_1, x_2, \dots, x_T)$  and  $U = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T)$ . Using the fact that  $\text{vec}(AB) = (B' \otimes I)\text{vec}(A)$ , where  $\otimes$  stands for the Kronecker product, the following univariate model can be written

$$\begin{aligned}\text{vec}(Z) &= (X' \otimes I_3) * \text{vec}(F) + \text{vec}(U) \\ z &= (X' \otimes I_3) * \text{vec}(F) + u\end{aligned}\tag{A.5}$$

$$\text{where Cov}(u) = \Omega = \text{diag}(Q_1, Q_2, \dots, Q_T).$$

Using the multivariate normal prior  $N(\underline{\mu}_{\text{vec}(F)}, \underline{\Sigma}_{\text{vec}(F)})$  in (11) the posterior distribution is Gaussian,  $\text{vec}(F)|z \sim N(\bar{\mu}_{\text{vec}(F)}, \bar{\Sigma}_{\text{vec}(F)})$ , with the following parameters

$$\begin{aligned}\bar{\Sigma}_{\text{vec}(F)} &= \left( \underline{\Sigma}_{\text{vec}(F)}^{-1} + (X' \otimes I_3)' \Omega^{-1} (X' \otimes I_3) \right)^{-1} \\ \bar{\mu}_{\text{vec}(F)} &= \bar{\Sigma}_{\text{vec}(F)} \left( \underline{\Sigma}_{\text{vec}(F)}^{-1} \underline{\mu}_{\text{vec}(F)} + (X' \otimes I_3)' \Omega^{-1} (X' \otimes I_3) \text{vec}(F)_{OLS} \right)\end{aligned}$$

$$\text{where } \text{vec}(F)_{OLS} = ((X' \otimes I_3)' \Omega^{-1} (X' \otimes I_3))^{-1} (X' \otimes I_3)' \Omega^{-1} z.\tag{A.6}$$

## A.2 Sampling of covariance matrix in the measurement equation, $R$

First, I rewrite the measurement equation

$$y_t = H f_t + \eta_t.\tag{A.7}$$

Since  $R$  is the covariance matrix of  $\eta_t$ ,  $R$  can be sampled from the inverse-Wishart distribution  $IW(\bar{\nu}_R, \bar{S}_R)$  with the degrees of freedom parameter  $\bar{\nu}_R = \underline{\nu}_R + T$  and with the shape parameter  $\bar{S}_R = \underline{S}_R + \sum_{t=1}^T (y_t - H f_t)(y_t - H f_t)'$ .

### A.3 Sampling factors, $f_t$

Conditional on the remaining model parameters, drawing  $f_{0:T}$  can be implemented using standard Bayesian inference. This constitutes running the Kalman filter first and running a simulation smoother using the filtered values for drawing smoothed states as in Carter and Kohn (1994) and Frühwirth-Schnatter (1994). In the first step, start the recursion for  $t = 1, \dots, T$

$$\begin{aligned}
f_{t|t-1} &= \mu_t + F f_{t-1|t-1} \\
P_{t|t-1} &= F P_{t-1|t-1} F' + Q_t \\
\eta_{t|t-1} &= y_t - H f_{t|t-1} \\
\zeta_{t|t-1} &= H P_{t|t-1} H' + R \\
K_t &= P_{t|t-1} H' \zeta_{t|t-1}' \\
f_{t|t} &= f_{t|t-1} + K_t \eta_{t|t-1} \\
P_{t|t} &= P_{t|t-1} - K_t H' \zeta_{t|t-1}',
\end{aligned} \tag{A.8}$$

and store  $f_{t|t}$  and  $P_{t|t}$ . For  $f_{0|0}$  and  $P_{0|0}$ , notice that, in the steady state  $E(f) = E(\mu) + F E(f)$  and  $E(P) = F E(P) F' + E(Q)$  leading to  $E(f) = (I_3 - F)^{-1} E(\mu)$  and  $\text{vec}(E(P)) = (I_3 - F \otimes F)^{-1} \text{vec}(Q)$ . Moreover, since the initial state is the state before observing data by definition,  $\mu$  and  $Q$  can be sampled from the base prior  $G_0$ .

The last filtered state  $f_{T|T}$  and its covariance matrix  $P_{T|T}$  correspond to the smoothed estimates of the mean and the covariance matrix of the factors for period  $T$ . Having stored all the filtered values, simulation smoother involves the following backward recursions for  $t = T - 1, \dots, 1$

$$\begin{aligned}
\eta_{t+1|t}^* &= f_{t+1} - \mu_{t+1} - F f_{t|t} \\
\zeta_{t+1|t}^* &= F P_{t|t} F' + Q_{t+1} \\
f_{t|t, f_{t+1}} &= f_{t|t} + P_{t|t} F' \zeta_{t+1|t}^{*-1} \eta_{t+1|t}^* \\
P_{t|t, P_{t+1}} &= P_{t|t} - P_{t|t} F' \zeta_{t+1|t}^{*-1} F P_{t|t}.
\end{aligned} \tag{A.9}$$

Intuitively, the simulation smoother updates the states using the same principle as in the Kalman filter, where at each step filtered values are updated using the smoothed values obtained from backward recursion. For updating the initial states, using the state equation  $f_{0|t,f_1} = F^{-1}(f_1 - \mu_1)$  and  $P_{0|t,P_1} = F^{-1}(P_1 + Q_1)F'^{-1}$  can be written for the first observation. Given the mean  $f_{t|t,f_{t+1}}$  and the covariance matrix  $P_{t|t,P_{t+1}}$ , the states can be sampled from  $f_t \sim N(f_{t|t,f_{t+1}}, P_{t|t,P_{t+1}})$  for  $t = 0, \dots, T$ .

#### A.4 Sampling decay parameter, $\lambda$

To sample  $\lambda$  from its conditional posterior distribution we can use (2). Unfortunately, the resulting distribution does not have a standard form, and thus, Gibbs sampling cannot be implemented. Therefore, I setup a Metropolis-Hastings (MH) (Metropolis et al., 1953; Hastings, 1970) step for sampling  $\lambda$ . As the candidate density an obvious alternative is to use a  $t$ -distribution with the maximum likelihood estimates of  $\lambda$  and its variance. However, this involves non-linear optimization procedures complicating the estimation process further. Instead, given the existing evidence on the decay parameter I use a  $t$ -distribution with a degrees of freedom 10 and with mean 0.07 and standard deviation 0.01. This candidate distribution puts a high probability mass on the region covering both 0.0609 and 0.077, the two values that are used for fixing the decay parameter in many applications of the DNS model. The resulting MH sampler performs quite well with a high acceptance rate for the model.

## References

- Ang, A., G. Bekaert, and M. Wei, 2008, The term structure of real rates and expected inflation, *Journal of Finance* 63(2), 797–849.
- Antoniak, C. E., 1974, Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems, *The Annals of Statistics* 2(6), 1152–1174.
- Bianchi, F., H. Mumtaz, and P. Surico, 2009, Dynamics of the term structure of uk interest rates, March, Bank of England.
- Blackwell, D., and J. MacQueen, 1973, Ferguson distributions via Polya urn schemes, *The Annals of Statistics* 1, 353–355.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31(3), 307–327.
- Caron, F., M. Davy, A. Doucet, E. Duflos, and P. Vanheeghe, 2008, Bayesian inference for linear dynamic models with dirichlet process mixtures, *IEEE Transactions on Signal Processing* 56(1), 71–84.
- Carter, C. K., and R Kohn, 1994, On gibbs sampling for state space models, *Biometrika* 81(3), 541–553.
- Chib, S., and B. H. Hamilton, 2002, Semiparametric bayes analysis of longitudinal data treatment models, *Journal of Econometrics* 110(1), 67–89.
- Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch, 2011, The affine arbitrage-free class of nelson-siegel term structure models, *Journal of Econometrics* 164(1), 4–20.
- Christensen, J. H. E., J. A. Lopez, and G. D. Rudebusch, 2010, Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields, *Journal of Money, Credit and Banking* 42(s1), 143–178.
- Conley, T. G., C. B. Hansen, R. E. McCulloch, and P. E. Rossi, 2008, A semi-parametric bayesian approach to the instrumental variable problem, *Journal of Econometrics* 144(1), 276–305.
- Cox, J. C., Jo. E. Jr. Ingersoll, and S. A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53(2), 385–407.
- Dai, Q., and K. J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55(5), 1943–1978.
- Dai, Q., and Kenneth J. S., 2002, Expectation puzzles, time-varying risk premia, and affine models of the term structure, *Journal of Financial Economics* 63(3), 415–441.

- De Pooter, M., 2007, Examining the nelson-siegel class of term structure models, June, Tinbergen Institute.
- Diebold, F. X., and C. Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* 130(2), 337–364.
- Diebold, F. X., G. D. Rudebusch, and B. S. Aruoba, 2006, The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics* 131(1-2), 309–338.
- Doucet, A., and C. Andrieu, 2001, Iterative algorithms for state estimation of jump markov linear systems, *IEEE Transactions on Signal Processing* 49(6), 1216–1227.
- Duffee, G. R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57(1), 405–443.
- Duffie, D., and R. Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6(4), 379–406.
- Escobar, M. D., 1994, Estimating Normal Means With a Dirichlet Process Prior, *Journal of the American Statistical Association* 89(425), 268–277.
- Escobar, M. D., and M. West, 1995, Bayesian Density Estimation and Inference Using Mixtures, *Journal of the American Statistical Association* 90(430), 577–588.
- Fama, E. F., and R. R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77(4), 680–92.
- Ferguson, T. S., 1973, A Bayesian Analysis of Some Nonparametric Problems, *The Annals of Statistics* 1(2), 209–230.
- Friedman, M., 1984, Lessons from the 1979-82 monetary policy experiment, *American Economic Review* 74(2), 397–400.
- Frühwirth-Schnatter, S., 1994, Data augmentation and dynamic linear models, *Journal of Time Series Analysis* 15, 183–202.
- Geman, S., and D. Geman, 1984, Stochastic relaxations, gibbs distributions, and the bayesian restoration of images, *IEEE Transaction on Pattern Analysis and Machine Intelligence* 6, 721–741.
- Geweke, J., and G. Amisano, 2010, Comparing and evaluating bayesian predictive distributions of asset returns, *International Journal of Forecasting* 26(2), 216–230.
- Griffin, J. E., and M. F. J. Steel, 2004, Semiparametric bayesian inference for stochastic frontier models, *Journal of Econometrics* 123(1), 121–152.
- Gurkaynak, R. S., B. Sack, and J. H. Wright, 2007, The u.s. treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54(8), 2291–2304.



- H., Keisuke, 2002, Semiparametric bayesian inference in autoregressive panel data models, *Econometrica* 70(2), 781–799.
- Hastings, W. K., 1970, Monte carlo sampling using markov chains and their applications, *Biometrika* 57, 97–10.
- Hautsch, N., and F. Yang, 2010, Bayesian inference in a stochastic volatility nelson-siegel model, *Computational Statistics & Data Analysis* In Press, Corrected Proof, –.
- Hautsch, N., and Y. Ou, 2009, Analyzing interest rate risk: Stochastic volatility in the term structure of government bond yields, Center for Financial Studies.
- Hull, J., and A. White, 1990, Pricing interest-rate-derivative securities, *Review of Financial Studies* 3(4), 573–92.
- Jacquier, E., N. G. Polson, and P. E. Rossi, 2004, Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* 122(1), 185–212.
- Jensen, M. J., 2004, Semiparametric bayesian inference of long-memory stochastic volatility models, *Journal of Time Series Analysis* 25(6), 895–922.
- Jensen, M. J., and J. M. Maheu, 2010, Bayesian semiparametric stochastic volatility modeling, *Journal of Econometrics* 157(2), 306–316.
- Junker, M., A. Szimayer, and N. Wagner, 2006, Nonlinear term structure dependence: Copula functions, empirics, and risk implications, *Journal of Banking & Finance* 30(4), 1171–1199.
- Kim, C.-J., and C. R. Nelson, 1999, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, 1 ed., vol. 1 ).
- Koopman, S. J., M. I. P. Mallee, and M. Van der Wel, 2010, Analyzing the term structure of interest rates using the dynamic nelson-siegel model with time-varying parameters, *Journal of Business & Economic Statistics* 28(3), 329–343.
- Litterman, R. B., and J. Scheinkemann, 1991, Common factors affecting bond returns, *The Journal of Fixed Income* 1, 54–61.
- MacEachern, S. N., and P. Müller, 1998, Estimating Mixture of Dirichlet Process Models, *Journal of Computational and Graphical Statistics* 7(2), 223–238.
- McCulloch, J. H., 1975, The tax-adjusted yield curve, *Journal of Finance* 30(3), 811–30.
- Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, 1953, Equations of state calculations by fast computing machines, *Journal of Chemical Physics* 21, 1087–1092.

- Neal, R.M., 2000, Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics* 249–265.
- Nelson, C. R., and A. F. Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business* 60(4), 473–89.
- Nyholm, K., and R. Rebonato, 2008, Long-horizon yield curve projections: comparison of semi-parametric and parametric approaches, *Applied Financial Economics* 18(20), 1597–1611.
- Omori, Y., S. Chib, N. Shephard, and J. Nakajima, 2007, Stochastic volatility with leverage: Fast and efficient likelihood inference, *Journal of Econometrics* 140(2), 425–449.
- Orphanides, A., and D. H. Kim, 2005, Term structure estimation with survey data on interest rate forecasts, November, Society for Computational Economics.
- Patton, A. J., and D. H. Oh, 2011, Modelling dependence in high dimensions with factor copulas, Manuscript.
- Ritter, C., and M. A. Tanner, 1992, Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler, *Journal of the American Statistical Association* 87(419), 861–868.
- Sethuraman, J., 1994, A constructive definition of Dirichlet priors, *Statistica Sinica* 4, 639–650.
- Siegel, A. F., and C. R. Nelson, 1988, Long-term behavior of yield curves, *Journal of Financial and Quantitative Analysis* 23(01), 105–110.
- Svensson, L. E. O., 1994, Estimating and interpreting forward interest rates: Sweden 1992 - 1994, September, National Bureau of Economic Research, Inc.
- Tanner, M. A., and W. H. Wong, 1987, The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association* 82, 528–550.
- Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5(2), 177–188.
- West, M., P. Müller, and M. D. Escobar, 1994, Hierarchical priors and mixture models with applications in regression and density estimation, *Aspects of Uncertainty* 363–386.
- Yu, W.-C., and E. Zivot, 2011, Forecasting the term structures of treasury and corporate yields using dynamic nelson-siegel models, *International Journal of Forecasting* 27(2), 579–591.

Table 1: Posterior results for the DNS-N, DNS-DPMP and DNS-DPM models

	<u>DNS-N</u>			<u>DNS-DPMP</u>			<u>DNS-DPM</u>		
	Level	Slope	Curvature	Level	Slope	Curvature	Level	Slope	Curvature
$\mu_f$	0.106 (0.031)	-0.075 (0.034)	-0.062 (0.039)	0.093 (0.026)	-0.026 (0.025)	-0.084 (0.035)			
$\phi_f$	0.985 (0.004)	0.959 (0.013)	0.919 (0.020)	0.987 (0.003)	0.979 (0.008)	0.939 (0.019)	0.986 (0.003)	0.978 (0.008)	0.947 (0.018)
$\sigma_f^2$	0.114 (0.008)	0.337 (0.023)	0.684 (0.057)						
$\rho_{1j}$	1	0.097 (0.049)	0.079 (0.056)						
$\rho_{2j}$	0.097 (0.049)	1	0.069 (0.052)						
$\rho_{3j}$	0.079 (0.056)	0.069 (0.052)	1						
$\alpha$					0.92 (0.50)			0.94 (0.47)	
# of clusters					5.77 (1.99)			5.84 (1.65)	
<u>Marginal likelihood values for Selected maturities</u>									
3 months		-338.99			-288.35			-279.86	
12 months		-254.70			-188.55			-166.18	
24 months		-235.45			-161.16			-139.55	
60 months		-183.24			-115.14			-86.77	
120 months		-169.81			-146.84			-129.34	

*Note:* DNS is the dynamic Nelson-Siegel model proposed by Diebold and Li (2006). DNS-N is the DNS model with Gaussian innovations, while DNS-DPM(P) is the model with DPM innovations, where both the mean and the variance (only the variance) of the innovations have a DP prior.  $\mu_f$  for  $f = 1, 2, 3$  is the unconditional mean of the factor errors (intercept of the factors) and  $\phi_f$  is the first-order autoregressive coefficients of the factors, i.e diagonal elements of the autoregressive coefficient matrix.  $\sigma_f^2$  is the variance of the factor error terms, whereas  $\rho_{i,j}$  denotes the correlation of the error term of the  $i^{th}$  factor for  $i = 1, 2, 3$  with the  $j^{th}$  factor error term  $j \in \{1, 2, 3, \setminus i\}$ .  $\alpha$  is the precision parameter of the DP. The values indicate the posterior means of the parameters with the posterior standard deviations in parenthesis. Marginal likelihood values are computed using (22). For the estimation of the models I use the sample period from January 1970 through December 2009. 60,000 simulations from the posterior distribution are used for inference. The first 10,000 draws are discarded as burn-in sample and the remaining 50,000 draws, where I kept every 10<sup>th</sup> draw, are used for posterior inference in the models. The convergence of the sampler is checked using statistical and visual inspection and in all model specifications convergence is assured.

Table 2: Posterior results (in basis points) for the variances of the measurement errors for the competing models

	<u>DNS-N</u>	<u>DNS-DPMP</u>	<u>DNS-DPM</u>	<u>DNS-DPM-<math>\lambda</math></u>
M3	12.58 (0.94)	12.20 (0.93)	12.18 (0.90)	12.23 (1.24)
M6	2.19 (0.21)	2.01 (0.20)	2.00 (0.19)	1.87 (0.48)
M9	0.82 (0.09)	0.74 (0.08)	0.74 (0.08)	0.67 (0.24)
M12	0.55 (0.05)	0.55 (0.05)	0.56 (0.05)	0.59 (0.06)
M15	0.72 (0.06)	0.75 (0.06)	0.75 (0.06)	0.77 (0.08)
M18	0.65 (0.05)	0.67 (0.05)	0.67 (0.05)	0.68 (0.06)
M21	0.63 (0.05)	0.64 (0.05)	0.64 (0.05)	0.63 (0.05)
M24	0.41 (0.03)	0.41 (0.03)	0.41 (0.03)	0.40 (0.03)
M27	0.31 (0.02)	0.31 (0.02)	0.31 (0.02)	0.29 (0.03)
M30	0.27 (0.02)	0.27 (0.02)	0.27 (0.02)	0.26 (0.03)
M33	0.22 (0.02)	0.22 (0.02)	0.22 (0.02)	0.22 (0.02)
M36	0.20 (0.02)	0.20 (0.02)	0.19 (0.02)	0.20 (0.02)
M39	0.25 (0.02)	0.25 (0.02)	0.25 (0.02)	0.25 (0.02)
M42	0.24 (0.02)	0.24 (0.02)	0.24 (0.02)	0.25 (0.02)
M45	0.28 (0.02)	0.28 (0.02)	0.28 (0.02)	0.28 (0.02)
M48	0.35 (0.03)	0.35 (0.03)	0.35 (0.03)	0.35 (0.03)
M60	0.41 (0.03)	0.40 (0.03)	0.40 (0.03)	0.37 (0.05)
M72	0.74 (0.06)	0.73 (0.06)	0.72 (0.06)	0.68 (0.07)
M84	0.75 (0.06)	0.73 (0.06)	0.73 (0.06)	0.85 (0.18)
M96	0.61 (0.07)	0.61 (0.07)	0.61 (0.07)	0.96 (0.42)
M108	1.84 (0.16)	1.89 (0.17)	1.89 (0.16)	2.46 (0.70)
M120	4.46 (0.34)	4.52 (0.35)	4.52 (0.34)	5.42 (1.13)

*Note:*  $M\tau$  indicates the yields with maturity  $\tau$  months. The results show the posterior means and posterior standard deviations (in parenthesis) of the variance of the measurement errors in (2), i.e. the diagonal elements of  $R$ . DNS-DPM- $\lambda$  stands for the model where the decay rate parameter  $\lambda$  is also estimated along with the other model parameters. See Table 1 for further details.

Table 3: Posterior predictive p-values ( $ppp$ ) of the models with skewness and kurtosis of the yields as criterion functions

	DNS-N		DNS-DPMP		DNS-DPM		DNS-DPM- $\lambda$	
	Skewness $ppp$	Kurtosis $ppp$	Skewness $ppp$	Kurtosis $ppp$	Skewness $ppp$	Kurtosis $ppp$	Skewness $ppp$	Kurtosis $ppp$
M3	0.04	0.45	0.02	0.23	0.03	0.32	0.09	0.22
M6	0.24	0.42	0.41	0.31	0.45	0.40	0.49	0.33
M9	0.05	0.23	0.33	0.46	0.31	0.43	0.30	0.48
M12	0.06	0.42	0.36	0.36	0.34	0.42	0.38	0.43
M15	0.04	0.30	0.29	0.21	0.30	0.26	0.34	0.28
M18	0.02	0.47	0.20	0.32	0.22	0.36	0.26	0.39
M21	0.00	0.27	0.14	0.49	0.15	0.42	0.19	0.44
M24	0.04	0.40	0.26	0.41	0.29	0.44	0.33	0.47
M27	0.10	0.27	0.41	0.23	0.43	0.23	0.47	0.25
M30	0.09	0.28	0.38	0.23	0.41	0.23	0.43	0.24
M33	0.10	0.32	0.39	0.26	0.43	0.26	0.43	0.26
M36	0.06	0.47	0.33	0.36	0.37	0.35	0.37	0.35
M39	0.04	0.48	0.25	0.40	0.29	0.38	0.29	0.38
M42	0.02	0.41	0.19	0.45	0.23	0.42	0.23	0.43
M45	0.01	0.49	0.17	0.40	0.20	0.37	0.20	0.38
M48	0.02	0.35	0.21	0.29	0.25	0.27	0.25	0.27
M60	0.09	0.31	0.36	0.27	0.41	0.24	0.42	0.26
M72	0.01	0.14	0.13	0.17	0.16	0.13	0.21	0.18
M84	0.10	0.36	0.38	0.50	0.43	0.44	0.54	0.45
M96	0.13	0.20	0.43	0.36	0.47	0.42	0.38	0.32
M108	0.08	0.03	0.34	0.13	0.37	0.17	0.56	0.12
M120	0.29	0.00	0.13	0.02	0.13	0.03	0.08	0.03

*Note:* M $\tau$  indicates the yields with maturity  $\tau$  months. DNS-DPM- $\lambda$  stands for the model, where the decay rate parameter  $\lambda$  is also estimated along with the other model parameters. The  $ppp$  values are computed using (23) with the objective functions of sample skewness and kurtosis. The  $ppp$  values correspond to the significance of the null hypothesis of the criterion functions' equality to that of the data. A  $ppp$  value closer to 0.5 indicates the better conformity of the model to the observed data. See Table 1 for further details.

Table 4: Posterior results for the DNS-DPM models with different hyperparameter settings

	Prior 2	Prior 3	Prior 4	Prior 5	Prior 6	Prior 7
$\phi_{11}$	0.987 (0.003)	0.987 (0.003)	0.985 (0.004)	0.985 (0.004)	0.986 (0.003)	0.986 (0.003)
$\phi_{22}$	0.978 (0.008)	0.977 (0.008)	0.978 (0.008)	0.978 (0.008)	0.978 (0.008)	0.978 (0.008)
$\phi_{33}$	0.947 (0.018)	0.946 (0.018)	0.937 (0.018)	0.932 (0.018)	0.937 (0.018)	0.932 (0.019)
$\alpha$	0.97 (0.50)	0.98 (0.50)	0.77 (0.46)	0.64 (0.43)	0.75 (0.46)	0.57 (0.41)
# of clusters	6.01 (1.84)	6.05 (1.92)	4.79 (1.75)	4.04 (1.69)	4.69 (1.82)	3.64 (1.61)

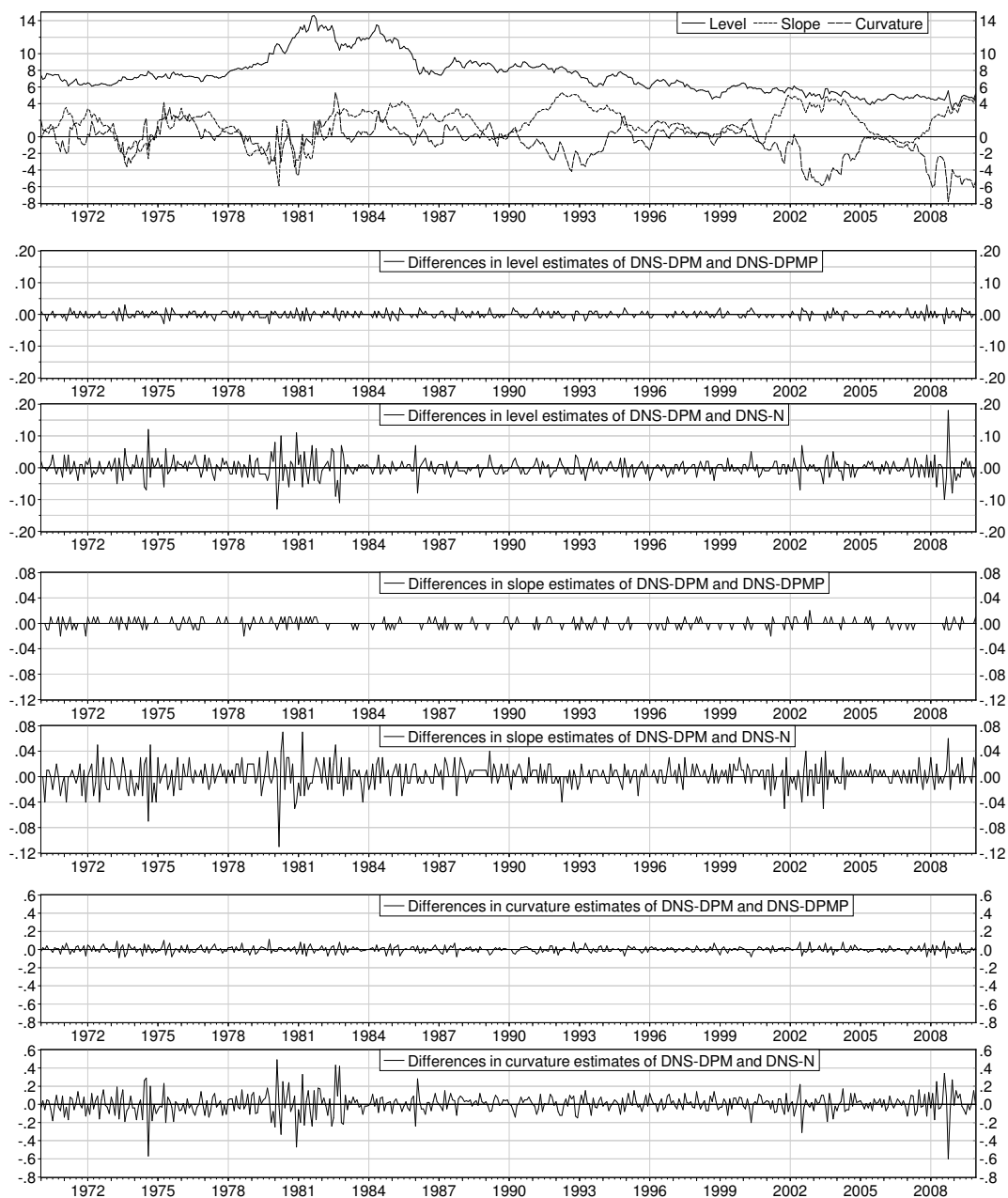
*Note:* The values indicate the posterior means of the parameters with the posterior standard deviations in parenthesis. See Table 1 for further details.

Table 5: Posterior results for the DNS-DPM- $\lambda$  model

	Level	Slope	Curvature
$\phi_{ii}$	0.986 (0.003)	0.977 (0.009)	0.952 (0.019)
$\lambda$		0.069 (0.011)	
$\alpha$		1.04 (0.50)	
# of clusters		6.41 (1.78)	

*Note:* DNS-DPM- $\lambda$  stands for the model, where the decay rate parameter  $\lambda$  is also estimated along with other model parameters. The values indicate the posterior means of the parameters with the posterior standard deviations in parenthesis. See Table 1 for further details.

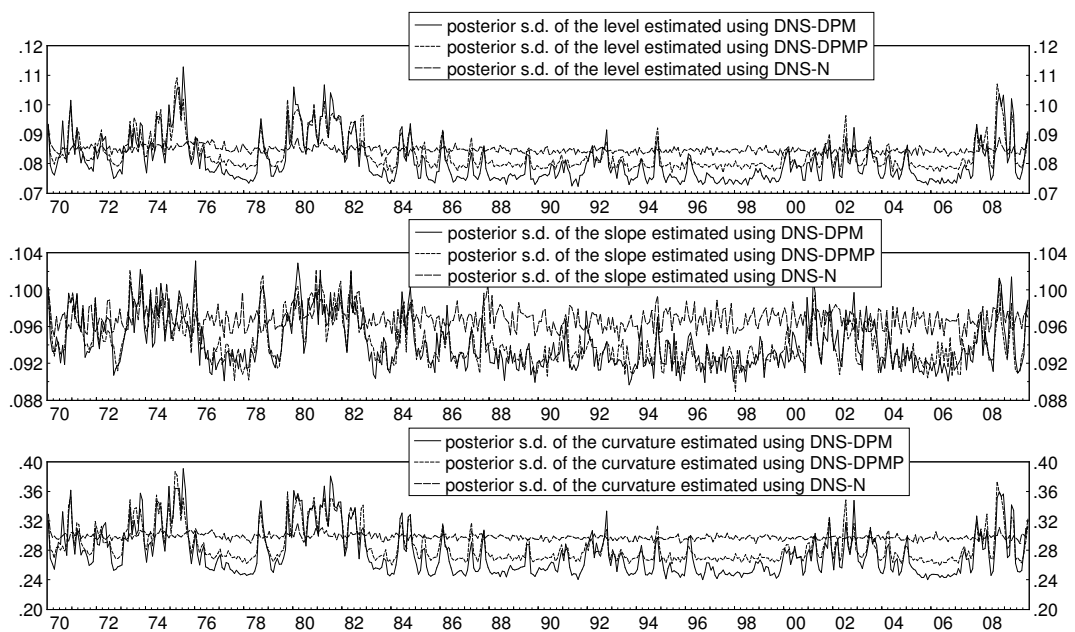
Figure 1: Posterior means of the factors for the DNS-DPM model and differences with those for the DNS-DPMP and DNS-N models



*Note:* DNS is the dynamic Nelson-Siegel model proposed by Diebold and Li (2006). DNS-N is the DNS model with Gaussian innovations, while DNS-DPM(P) is the model with DPM innovations, where both the mean and the variance (only the variance) of the innovations have a DP prior. In the first graph, the solid, the dotted, and the dashed lines represent posterior means of the level, slope, and curvature factors, respectively, estimated using the DNS-DPM model. In the remaining graphs the solid lines correspond to the differences of factor estimates (posterior means) of the competing models.

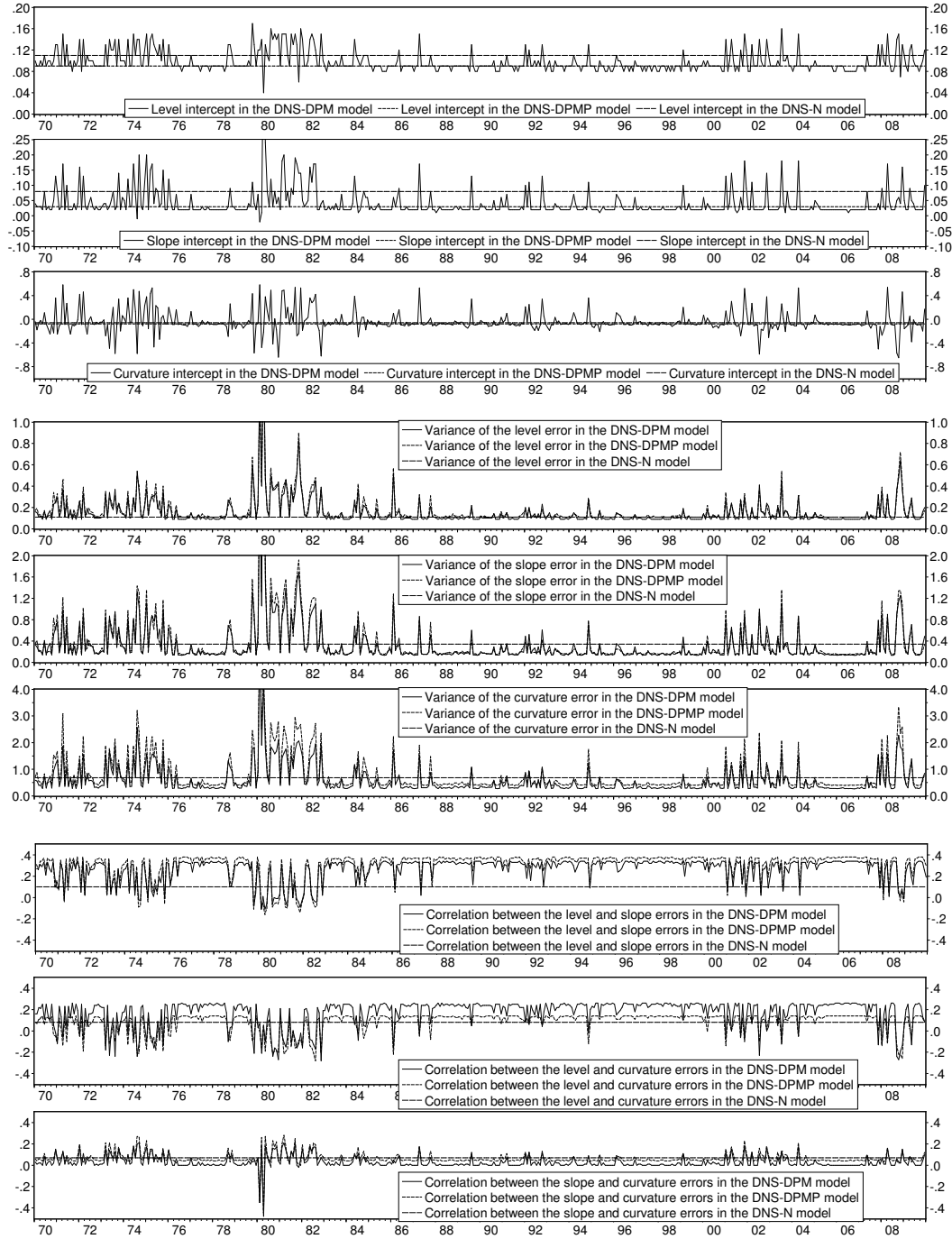


Figure 2: Posterior standard deviations of the factors for the DNS-DPM, DNS-DPMP and DNS-N models



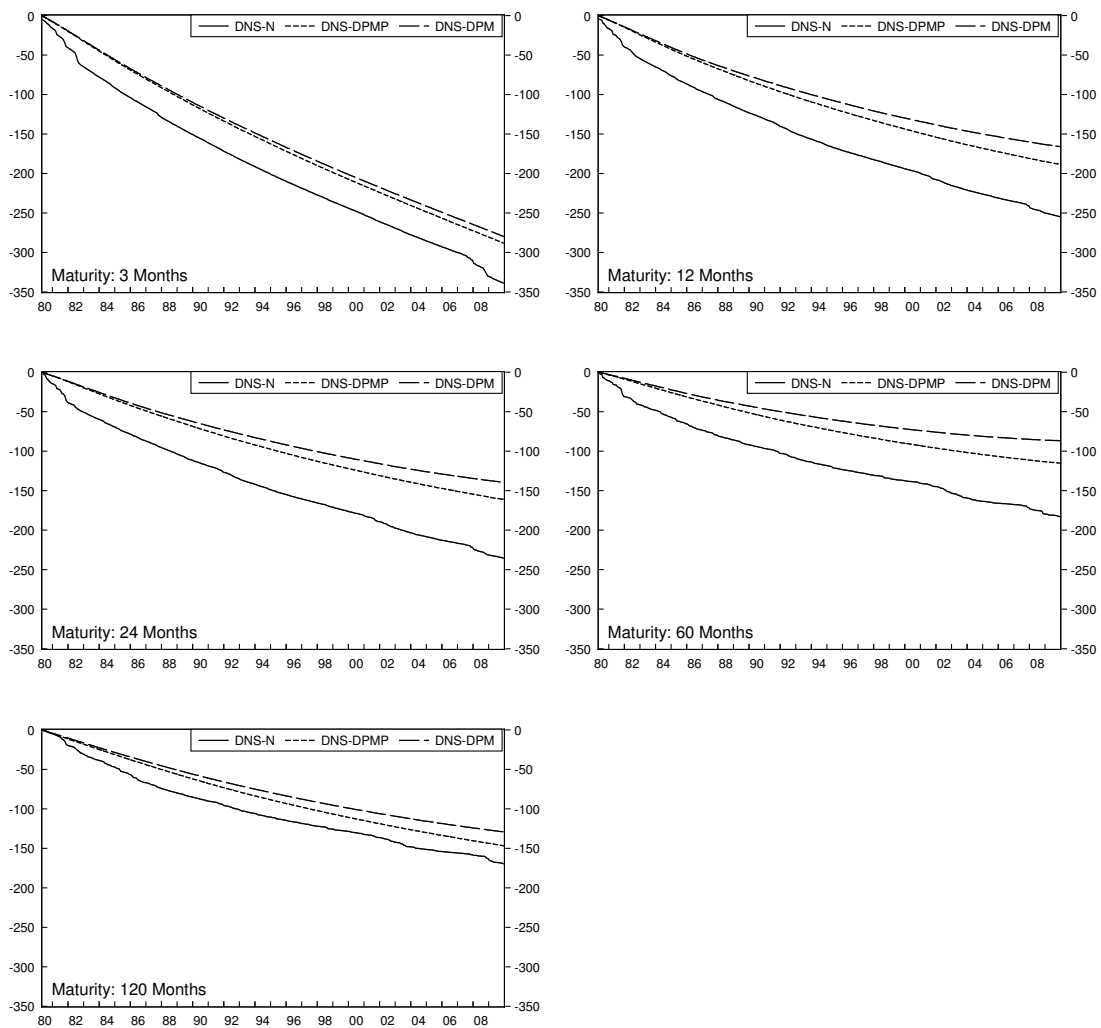
*Note:* The solid, the dotted, and the dashed lines represent posterior standard deviations of the factors estimated using DNS-DPM, DNS-DPMP, and DNS-N models, respectively. The first graph displays the results for the level factors, whereas the second and the third graph display the results for the slope and curvature factors, respectively. See Figure 1 for further details.

Figure 3: Posterior means of the parameters for the DNS-DPM, DNS-DPMP and DNS-N models



*Note:* The solid lines correspond to the estimates of the time varying parameters due to the DPM structure which are not displayed in Table 1. The dotted lines represent the corresponding estimates using DNS-N model displayed also in Table 1. See Table 1 for further details.

Figure 4: Evolution of the marginal likelihood values of DNS-DPM, DNS-DPMP and DNS-N models for selected maturities over time



*Note:* The solid lines correspond to the evolution of the marginal likelihood value of the DNS-N model. The dotted lines represent the corresponding values for DNS-DPMP model and the dashed lines represent the corresponding values for DNS-DPM model. See Table 1 for further details.