

Modeling the Density of US Yield Curve using Bayesian Semiparametric Dynamic Nelson-Siegel Model *

Cem Çakmaklı[†]

*Department of Economics
Koc University*

September 2016

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 an infinite Markov mixture model. The semiparametric structure of the factors enables us to capture various forms of non-normalities including fat tails, skewness and nonlinear dependence between factors using a unified approach. The potential of the proposed framework is examined using US bond yields data. 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. We document that modeling parameter changes in a discrete manner as opposed to the continuous changes as in stochastic volatility models increases the model fit as well as forecasting performance of the semiparametric model. This is due to fact that discrete changes suit the typical low frequency monthly bond yields data better.

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

JEL Classification: C14, C33, C38, G12

*We thank to Richard Paap and Dick van Dijk for valuable comments and suggestions in various phases of this research. We thank to Peter Boswijk, Marco Del Negro, Domenico Giannone, Jim Griffin, Lennart Hoogerheide, Mark Jensen, Maria Kalli, John Maheu, Hashem Pesaran, Simon Potter and Rodney Strachan

[†]Koc University, Department of Economics, Rumelifeneri Yolu, 34450, Saryer, Istanbul, Turkey, e-mail: ccakmakli@ku.edu.tr.

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, 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, modeling the density of bond yields does not reach a consensus yet. Often, error distributions of the factors are assumed to be Gaussian in order to facilitate the econometric inference. One option to extend the standard approach is adding a stochastic volatility or a GARCH process for modeling the density of the yield curve. This has the advantage of the fact that volatility can take any value governed by the volatility process, and hence, it provides a quite a general framework for modeling the density of the yield curve using the volatility process. However, modeling volatility process by continuous changes as in the stochastic volatility or GARCH models is mostly suitable for high frequency financial datasets. Alternatively, regime switching models with Markov dependence structure also constitutes an important group of models where changes in parameters are discrete, which seems to be more suitable for monthly macro-financial datasets of treasury bond yields. In this approach, continuous changes are replaced by discrete changes together with a limited number of regimes. This implies that the values that parameters can take usually are confined to two or three depending on the pre-specified number of regimes. This restriction of the number of regimes precludes density estimation as an thoroughly efficient density estimation

requires a much flexible approach.

In this paper, we propose a Bayesian semiparametric alternative for the estimation of the density of the yield curve. Specifically, we specify the distribution of the underlying yield curve factors using an infinite (Markov) mixture that is proved to be very useful in many applications of econometrics, see for example Chib and Hamilton (2002); Hirano (2002); Griffin and Steel (2004); Jensen (2004); Conley *et al.* (2008); Jensen and Maheu (2010), among others. The proposed model combines the strength of both of the density modeling structures. While the model allows for discrete changes in the parameters as in the standard regime switching models, it also allows the parameters to take any values as in the stochastic volatility models using the infinite Markov mixture structure. The resulting model does not rely on the assumption of a specific distribution of the factors nor we presume a specific number of distributions for the mixture distribution of the factors. Instead, using the semiparametric structure we estimate the factor density, i.e. number of regimes, along with the factors, and thus, we indeed estimate the density of the factors and thereby the density of the yield curve.

Traditional approaches of modeling the dynamics of the yield curve mostly focus either on the no-arbitrage relations in equilibrium to model the yield curve¹ 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, 1988) provide a parsimonious three-component exponential approximation to the yield curve, which is extended to allow for time dependence between the components by Diebold and Li (2006). Diebold *et al.* (2006) further show that the resulting

¹See 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 Singleton (2002) and Duffee (2002), among others, impose some further restrictions on model parameters to alleviate these problems.

dynamic model can be cast in a dynamic factor model framework with three factors. The ‘dynamic’ Nelson-Siegel model (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.

The DNS 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, its use for estimating the density of the yield curve is relatively limited. Typically, the distribution of the factors and the yields are specified as Gaussian with constant parameters to facilitate the estimation. Many studies document evidence of deviations from the Gaussian distributions indicating the need for an explicit density estimation due to several reasons. An obvious reason is that for US bond yields the volatility before the Great Moderation² had been much higher than the succeeding period. Hautsch and Ou (2009), Koopman *et al.* (2010) Carriero *et al.* (2014), Mesters *et al.* (2014) for the U.S. and Bianchi *et al.* (2009) for the U.K., among others, document the variation in the volatility of the yields and/or factors even for the Great Moderation period after mid 1980’s. These studies mostly employ stochastic volatility (Jacquier *et al.*, 2004; Omori *et al.*, 2007) or GARCH (Bollerslev, 1986) type model structures to capture the fat tails of the yield distribution and the time variation in the volatility.

However, these model structures with continuous parameter changes are conventionally used to fit the high frequency financial data at daily or even at intra-daily frequency. When the focus is on bond yields at lower frequency the suitability of these approaches are not certain. One reason is that the bond yields are affected

²Persistent decline in the volatility of many US macroeconomic and financial indicators after mid-1984 is referred as Great Moderation.

mostly by policy changes that are discrete in nature. Yields representing the expectations of the agents about the future state of the economy adapt the policy changes rapidly indicating a discrete change especially at monthly frequency. For the US, we observe such phenomena especially at the beginning of 1980s when aberrant values and quite irregular dynamics of bond yields were observed due to the monetary policy experiments of Federal Reserve, see Friedman (1984) for details. Indeed, Xiang and Zhu (2013) and Hevia *et al.* (2015), among others, provide a regime switching DNS model that allows for a fixed number of regimes capturing the changes in business cycle conditions and monetary policy stance. In our specification, we do not restrict the number of regimes but we estimate the number of regimes along with model parameters. In this sense, our model captures both regime switching models with pre-specified number of regimes and models with continuous changes, i.e. allowing for changes of parameters in each time period, as special cases.

Several issues plague the estimation of the density of the yields. As the yield curve is constructed using a large cross section of the yields, departure from standard assumptions for each bond yield often poses overparametrization of the resulting model complicating econometric inference. One elegant and natural way to sidestep overparametrization is estimating the density of the series by 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.³ This is even more valid for the DNS model where a few key factors explain the bulk of the variation in the entire cross section of the yields.

The Bayesian semiparametric DNS model provides a unified approach that can take all features of bond yields into account by estimating the distribution of the dynamic factors along with other model parameters. In the empirical application

³Using a similar intuition Oh and Patton (2016) model the nonlinear dependence in high dimensional models using factor copulas where factors are modeled using non-Gaussian distributions.

using panel time series of US yield data over the period starting from 1970 until 2009, we show that the Bayesian semiparametric DNS model captures 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 peaks of the yields in early 1980s during monetary policy experimentation as outliers in separate clusters leading to more precise estimates of the parameters. This is thanks to the infinite mixture structure where new clusters capture these outliers endogenously by estimating the number of clusters rather than pre-specifying it. 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 even switch signs indicating a nonlinear pattern. Finally, we perform a sensitivity analysis regarding to estimation period 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. We discuss the Bayesian estimation methodology together with the prior and posterior specifications in Section 3, with full details being provided in A. Section 4 displays and discusses the results of the model using US data. 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, 1988) provide 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, 1988) specify the forward curve using an exponential approximation, which translates into the following specification of the yield curve with only three components as

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

The main advantage of this parsimonious specification is its flexibility to accommodate different forms of the yield curve successfully. The parameter λ in (1) governs the exponential decay rate determining the maximum loading of the third coefficient. In the classical specification of Nelson and Siegel (1987, 1988), the key parameters $(f_{1,t}, f_{2,t}, f_{3,t})'$ are interpreted as static coefficients to be estimated at a given point in time. Diebold and Li (2006) extend the model by interpreting these parameters as factors governing the evolution of the cross section of yields over time. The resulting dynamic factors, then, can be interpreted as level, slope and curvature of the yield curve factors,⁴ see Diebold and Li (2006) for a detailed discussion.

Given this interpretation of the Nelson-Siegel factors, the evolution of these factors shapes the evolution of the entire yield curve over time. Specifically, Diebold

⁴The interpretation of the yield curve factors as level, slope and curvature is due to the seminal paper of Litterman and Scheinkman (1991). They use principle component analysis to extract the three common factors from a large set of US treasury bond yields. By focusing on the relation of each factor to the underlying yields with specific maturities they show that these factors can be interpreted as level, slope and curvature of the yield curve.

et al. (2006) extend this framework by imposing a first order autocorrelation structure to these factors to model the dependence of the factors temporally. The resulting state-space representation of the model 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 takes the following (vector) autoregressive form

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

In the standard dynamic Nelson-Siegel model the error terms in both equations are distributed according to a multivariate Gaussian distribution with fixed parameters. The resulting model can be written in the compact notation as

$$\begin{aligned} y_t &= H f_t + \eta_t & \eta_t &\sim N(0, R) \\ f_t &= F f_{t-1} + \varepsilon_t & \varepsilon_t &\sim N(\theta) \text{ where } \theta = \{\mu, Q\} \end{aligned} \quad (4)$$

where η_t follows a multivariate Gaussian distribution with mean zero and the covariance matrix R . The error distribution of the factors, ε_t , also follows a multivariate Gaussian distribution (independent of η_t) conditional on the model parameters θ , where θ includes the mean⁵ and the variance of the factor error distribution.

⁵The constant term in the state equation is suppressed in the error distribution as the mean of the error distribution.

2.2 Continuous changes in model parameters: time varying intercept and stochastic volatility

In the standard case, θ involves constant parameters yielding an optimum point prediction in the least squares sense. Although Bayesian inference provides a density estimation as the parameters are stochastic, rigorous estimation of the density requires time-variation in the parameters specifically in the presence of time varying volatility and potential outliers. Following this, the model can be extended to allow for time varying parameters, θ_t , in the form of continuous changes imposed by a time varying intercept and stochastic volatility as follows,

$$\begin{aligned}\mu_{it} &= \mu_{i,t-1} + \eta_{i,t}^\mu & \eta_{i,t}^\mu &\sim N(0, \sigma_{\mu_i}^2) \\ h_{it} &= \phi_{h_i} h_{i,t-1} + \eta_{i,t}^h & \eta_{i,t}^h &\sim N(0, \sigma_{h_i}^2) \quad \text{for } i = 1, 2, 3.\end{aligned}\tag{5}$$

Consistent with time varying parameter models, we let the intercepts follow a random walk, see, for example, Primiceri (2005). For the volatility process, however, we also allow factor volatilities to follow a stationary process, following the evidence in Hautsch and Yang (2010). Our findings suggest that while the persistence of level factor volatility is high, the persistence of the remaining factor volatilities are relatively lower.⁶ Therefore, we fix the value of the autoregressive parameter for the volatility of the level factor at unity, imposing a random walk process but we estimate the remaining parameters by also allowing stationarity. The resulting model allows the intercept and the variance to change in a continuous manner leading these parameters to operate in an infinite dimensional parameter space smoothly. Still, we observe some sudden major changes in the yield data especially during 1979-1982 period where the rapid changes are as high as 4 percentage points. Interestingly, these jumps are common to all maturities indicating a maturity wide time specific shock. To further allow for these sudden changes, we extend the distribution of the

⁶Results are available upon request.

factor errors to follow a t distribution with the prior degrees of freedom of 20.⁷

2.3 Discrete changes in model parameters: infinite Markov mixture model

We start this section with a brief description of regime switching model with Markov dynamics, or in other terms, finite Markov mixture (fMM) model. We then show that in the limiting case fMM model becomes infinite Markov mixture (iMM) model. A fMM model consists of latent states⁸ where the number of the states, K , are determined prior to analysis. Let denote a state in period t as s_t where $s_t = 1, \dots, K$. Taking time dependence into account transitions between states follow a Markov process of first order, $\pi_{ij} = \Pr(s_t = j | s_{t-1} = i)$. In our specification, conditional on each of state, error terms follow a Gaussian distribution with parameters $\theta_t \equiv \theta_{s_t}$. This implies that the parameters can take K specific values depending on the type of the state in period t . In a typical analysis the transition probabilities from state i follow a Dirichlet distribution. Therefore, for the conjugate prior specification we also use a Dirichlet distribution of symmetric form with parameter γ/K assigning equal probability for each of transition reflecting a non-informative proper prior distribution. To obtain an iMM model let $K \rightarrow \infty$, to obtain a Dirichlet 'process' prior. One of the obstacles of this approach in the Markov mixture models content is that we assign a different Dirichlet process prior for transition from each state. Therefore there is no coupling across transitions out of different states since the transition probabilities are given independent priors. To introduce coupling, we can define a modified version of fMM and then let $K \rightarrow \infty$, as

$$\begin{aligned} \pi_i &\sim \text{Dirichlet}(\alpha\beta) \\ \beta &\sim \text{Dirichlet}(\gamma/K) \end{aligned} \tag{6}$$

⁷Different degrees of freedom parameters such as 10 and 15 give quite similar results.

⁸We use the the words 'regime' and 'state' interchangeably.

where π_i denotes transition probabilities out of state i . In this specification distribution of all transition probabilities share a common parameter, β , which itself follows a Dirichlet distribution. As $K \rightarrow \infty$ this hierarchical prior approaches to hierarchical Dirichlet process. This prior together with a standard likelihood yields the iMM model. The advantages of iMM model is threefold. First, as in the stochastic volatility case, the parameter values at t can take any values unlike fMM model. Second, iMM model spans all continuous distributions with probability one indicating that it is thoroughly a density estimation. Third, changes in the parameters follow a discrete manner suitable for the low frequency macro-finance data.

We now proceed with a more technical exposition of the iMM model. Let G_0 be a probability distribution function over some parameter space Θ such that $\theta \in \Theta$ and let α be a positive scalar. A distribution G is distributed with Dirichlet process $DP(\alpha, G_0)$ with base measure G_0 and precision parameter α 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 $Dir(\alpha G_0(\theta_1), \alpha G_0(\theta_2), \dots, \alpha G_0(\theta_J))$. Specifically

$$f(\theta_t | \theta_{-t}) | G = G = \frac{\alpha}{\alpha + T - 1} G_0(\Lambda) + \sum_{i=1, i \neq t}^T \frac{\delta^j(\theta_i)}{\alpha + T - 1}, \quad (7)$$

where $\delta^j(\theta_i)$ is the Dirac delta which takes the value of one if $\theta_i = \theta_j$ and zero otherwise. Dirichlet process spans the entire distribution of the discrete probability functions with probability one. This can be seen from the stick breaking representation of the DP, see Sethuraman (1994), with the following scheme. Suppose a part of $W_j\%$ of a stick with a unit initial length is being constantly broken j times where $W_j \sim Beta(1, \alpha)$. 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 one. This stick-breaking construction is denoted by $\pi \sim GEM(\alpha)$.

2.4 Estimating the density of the yield curve

Together with iMM model, we complete the set of competing models. The first model constitutes the standard dynamic Nelson model, denoted as DNS-N, as follows

$$\begin{aligned} y_t &= H f_t + \eta_t & \eta_t &\sim N(0, R) \\ f_t &= \mu + F f_{t-1} + \varepsilon_t & \varepsilon_t &\sim N(0, Q) \end{aligned} \quad (8)$$

The second model, denoted as DNS-SVt, involves stochastic volatility where logarithm of volatility is denoted as h_i for $i = 1, 2, 3$ together with time varying intercept. Further, the distribution of the error term follows a t distribution with degrees of freedom ν as follows

$$\begin{aligned} y_t &= H f_t + \eta_t & \eta_t &\sim N(0, R) \\ f_t &= \mu_{t-1} + F f_{t-1} + \varepsilon_t & \varepsilon_{i,t} &\sim t(0, \nu, \exp(h_{i,t})) \\ \mu_{i,t} &= \mu_{i,t-1} + \eta_{i,t}^\mu & \eta_{i,t}^\mu &\sim N(0, \sigma_{v_i}^2) \\ h_{it} &= \phi_{h_i} h_{i,t-1} + \eta_{i,t}^h & \eta_{i,t}^h &\sim N(0, \sigma_{h_i}^2) \quad \text{for } i = 1, 2, 3 \end{aligned} \quad (9)$$

The final model specifies a Bayesian semiparametric structure of infinite Markov mixture models for the yield curve factors as follows

$$\begin{aligned} y_t &= H f_t + \eta_t & \eta_t &\sim N(0, R) \\ f_t &= F f_{t-1} + \varepsilon_t & \varepsilon_t &\sim N(\theta_t) \\ \theta_t | G &\sim DP(\alpha, G_0) & G_0 &\sim DP(\gamma, H) \end{aligned} \quad (10)$$

In the iMM model framework, the base measure G_0 itself follows a $DP(\gamma, H)$ with parameters γ and H as (global) base measure in an hierarchical framework. A crucial feature of the hierarchical DP is that the stick breaking representation leads to the base measures described in a general framework as $G_0 = \sum_{k'=1}^{\infty} \beta_{k'} \delta_{\theta_{k'}}$ and $G_k = \sum_{k'=1}^{\infty} \pi_{kk'} \delta_{\theta_{k'}}$ where $\beta \sim \mathbf{GEM}(\gamma)$.

Without specifying the autoregressive coefficient matrix F and the covariance

matrix of the measurement error R the model is not complete. Following the evidence on the persistence of the yield curve factors, we keep F out of the set of the parameters that are subject to time variation. Additionally, we restrict the off-diagonal elements to be zero in line with the existing evidence.⁹

Another parameter that we keep fixed is the covariance matrix of the measurement error, R . In fact, specifying a semiparametric structure for the distribution of R is a straightforward extension. 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. Our strategy is to estimate the density of the yields in a parsimonious yet flexible way by exploiting the underlying factor structure using the fact that factors can explain the bulk of the variation in a large set of yields.

3 Bayesian inference

In this section we provide details on econometric inference of the Bayesian semiparametric dynamic Nelson-Siegel model. The joint posterior distribution of the parameters with IMM structure is constructed by the product of the hierarchical DP prior together with the likelihood. In Section 3.1, we discuss the prior distributions. In Section 3.2, we discuss the full conditional posterior distributions and the resulting simulation scheme. We discuss the posterior conditional distributions which do not follow the IMM distribution in A as these distributions are identical with the parameter distributions in the standard DNS model with Gaussian innova-

⁹One of the most prominent stylized facts of the yield curve models is the persistence of the factors and insignificance of the off-diagonal elements. In many applications of the dynamic Nelson-Siegel model estimates of the autoregressive coefficient matrix, F , indicate uniformly this persistence with diagonal values approximately above 0.9 together with very low values of the remaining elements regardless of the model specification or estimation technique. 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. We also estimate the model with a full autoregressive coefficient matrix specification. Results are qualitatively identical and available upon request.

tions. The Bayesian model selection procedures for model comparison are discussed in Section 3.3. The details on the econometric inference of the DNS-N and DNS-SVt model is provided in A.

3.1 Prior distributions

In this section we specify the prior distributions of the parameters starting with the prior distributions related to the hierarchical Dirichlet process. This includes the specification of the global base prior, H , and the specification of the hyperpriors for the precision parameters α and γ . For the global base prior, we use a normal-inverse Wishart (NIW) distribution as follows

$$H(\mu_0, a, \nu, S) = \text{NIW}(\mu_0, a, \nu, S) \mapsto \mu|\Sigma \sim N\left(\mu_0, \frac{\Sigma}{a}\right) \text{ and } \Sigma \sim IW(\nu, S), \quad (11)$$

where $E(\mu) = \mu_0$ and $E(\Sigma^{-1}) = \nu S^{-1}$. Moreover, the scale parameter, S , can be decomposed further into $S = \nu\psi I_3$, where I_n is the identity matrix of the dimension n . This implies that a priori we 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. We calibrate the values of hyperprior parameters using the parameter estimates from a vector autoregression with yield curve factors constructed using specific combinations of maturities, see for example (Diebold and Li, 2006).

For the precision parameters, α and γ , we use the hyperprior distributions of the Gamma form

$$f(\alpha) \propto \Gamma(a_\alpha, b_\alpha) \quad \text{and} \quad f(\gamma) \propto \Gamma(a_\gamma, b_\gamma). \quad (12)$$

We prefer a Gamma specification with greater support on small values of α and γ by specifying $a_\alpha = a_\gamma = 2$ and $b_\alpha = b_\gamma = 1$ to maintain a balance between the density estimation and model overfitting. Since large values of precision parameters imply

a large number of clusters, more flexible specifications could easily lead to model overfitting.

We proceed with the remaining model parameters that do not involve a DP prior. For the prior distribution of the autoregressive parameters in (3), we assume the standard multivariate normal distribution as

$$f(\text{vec}(F)) = N(\mathbf{0}_9, I_9), \quad (13)$$

where $\text{vec}(F)$ is the vec operator stacking each columns of F into a single column vector. $\mathbf{0}_n$ is a vector of zeros of dimension n .¹⁰

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

$$f(\sigma_n^2) \propto \sigma_n^{-2} \quad \text{for all } n = 1, 2, \dots, N, \quad (14)$$

see Geisser (1965), where σ_n^2 's are the diagonal elements of R and N is the number of the cross section 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, we set up an MCMC algorithm to sample from the full conditional posterior distributions. Specifically, we use Gibbs sampling together with data augmentation (see Geman and Geman, 1984; Tanner and Wong, 1987) to obtain posterior results.

¹⁰Unreported results of prior sensitivity analysis suggests that the likelihood information dominates prior information for F and thus, the conventional prior specifications as our specification are of limited or no importance for this parameter.

For the inference on IMM structure we utilize the beam sampler proposed by Van Gael *et al.* (2008) which proved to be very efficient and robust compared to competing sampling schemes such as pure Gibbs sampling. Intuitively, Beam sampler combines the slice sampler of Neal (2003) which limits the number of states considered at each time step to a finite number, with standard forward-backward algorithms given the finite number of states. The resulting algorithm exploits the tractability of the conditional distributions thanks to the finite number of states due to slice sampler. Note that in each iteration of the MCMC sampler whole trajectory of the states are sampled in a multi-move fashion. This is much efficient than single-move algorithms where state in period t is sampled conditional on the adjacent states. The resulting simulation scheme is as follows;

- Initialize the parameters by drawing θ_t from the base prior H and initialize remaining parameters. At step (m) of the iteration
 1. sample $f_{0:T}^{(m)}$ from $p(f_{0:T}|y_{1:T}, \theta_{1:T}^{(m-1)}, F^{(m-1)})$,
 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 $p(\theta_t|f_0^{(m-1)}, f_{0:T}^{(m-1)}, F^{(m)})$ for $t = 1, 2, \dots, T$ using the beam sampler.
- Repeat (1)-(4) M times.

Steps (1)-(3) 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 A. Here we focus on the beam sampler in step (4).

3.2.1 Sampling of θ_t

Sampling the parameters with iMM structure 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. Here we rely on the Gibbs sampler and consider the factors estimated in each iteration as observed data. This facilitates the inference and enables to conduct the efficient beam sampler in our framework.

The efficiency of the beam sampler stems from the fact that it makes use of auxiliary variables for slicing the distribution which in turn transform an infinite mixture structure into a finite one. Specifically let u_t be an auxiliary variable in period t and let π be the matrix of transition probabilities with π_{ij} in the ij^{th} entry. Each u_t is distributed as $\text{Uniform}(0, \pi_{s_{t-1}s_t})$ conditional on π and the states s_t and s_{t-1} , indicating the cluster dependence as $s_t = i$ if θ_t belongs to the cluster i in period t , from previous iteration of the sampler. Conditional on u_t , only the transitions which satisfy $\pi_{s_{t-1}s_t} \geq u_t$ will have a non-zero probability. Hence, the slicing of the distribution consists of the decision of whether a transition will occur or not given the states from the previous iteration of the sampler.

Given $u_{1:t}$, using the conditional distribution of the states, $p(s_t|u_{1:t}, f_{1:t})$, the distribution of the states are as follows

$$\begin{aligned}
 p(s_t|f_{1:t}, u_{1:t}) &\propto p(s_t, u_t, f_t|f_{1:t-1}, u_{1:t-1}) \\
 &= \sum_{s_{t-1}} p(f_t|s_t)p(u_t|s_t, s_{t-1}), p(s_t|s_{t-1})p(s_{t-1}|f_{1:t-1}, u_{1:t-1}) \quad (15) \\
 &= p(f_t|s_t) \sum_{s_{t-1}} \mathbb{I}(u_t < \pi_{s_t, s_{t-1}}) p(s_{t-1}|f_{1:t-1}, u_{1:t-1})
 \end{aligned}$$

where $\mathbb{I}(C)$ becomes 1 if condition C true and 0 otherwise. The key in this derivation is the transition from step two to three, where the auxiliary variable u_t truncates the infinite sum over s_{t-1} into a finite sum as $\mathbb{I}(u_t < \pi_{s_t, s_{t-1}})$ becomes 0 for some of the potential transitions. Finally, backward recursion can be implemented by first sampling from $p(s_T|f_{1:T}, u_{1:T})$ and sampling smoothed states using

$$p(s_t | s_{t+1}, f_{1:T}, u_{1:T}) \propto p(s_t | f_{1:t}, u_{1:t}) p(s_{t+1} | s_t, u_{t+1}).$$

The transition probabilities from state i to j , $i, j \in 1, \dots, K$ where K is the total number of distinct states, can be represented as $(\pi_{i1}, \pi_{i2}, \dots, \pi_{iK}, \sum_{j=K+1}^{\infty} \pi_{ij})$. Notice that the probability $\sum_{j=K+1}^{\infty} \pi_{ij}$ captures the remaining infinitely many states. Using the hierarchical Dirichlet process prior and the Gaussian likelihood, the posterior is distributed as $\text{Dirichlet}(\alpha\beta_1 + n_{i1}, \alpha\beta_2 + n_{i2}, \dots, \alpha\beta_K + n_{iK}, \alpha \sum_{j=K+1}^{\infty} \beta_j)$.

Sampling of β involves to generate another set of auxiliary variables, m_{ij} which are distributed as

$$p(m_{ij} = m | \beta, \alpha, s_{1:T}) \propto \mathbb{S}(n_{ij, m})(\alpha\beta_j)^m \quad (16)$$

where $S(\cdot)$ denotes Stirling numbers of the first kind. Then $(\beta_1, \beta_2, \dots, \beta_K, \sum_{j=K+1}^{\infty} \beta_j)$ are distributed with $\text{Dirichlet}(m_{.1}, m_{.2}, \dots, m_{.K}, \gamma)$ where $m_{.j} = \sum_{i=1}^K m_{ij}$.

The posterior distribution of α , $p(\alpha | K)$, is the product of the hyperprior in (12) and the likelihood function $p(K | \alpha)$ where K is the number of clusters. The resulting posterior is

$$p(\alpha | K) \propto p(\alpha) T! \alpha^K \frac{\Gamma(\alpha)}{\Gamma(\alpha + T)}, \quad (17)$$

see Antoniak (1974). West (1992) shows that the distribution of α conditional on the number of clusters can be cast as a mixture of two Gamma distributions with the mixing proportion distributed according to the Beta distribution. Hence, a Gamma prior for α as in (12) leads to a tractable posterior distribution of mixture of two Gamma distributions. Following this, a sample from α can be obtained by sequentially sampling first the mixing proportion from the Beta distribution and then α from the Gamma distribution.

3.3 Model selection

The Bayesian semiparametric dynamic Nelson-Siegel model encompasses the standard DNS model with Gaussian innovations as well as the DNS model with con-

tinuous parameter changes as specific cases. Hence, comparison of the standard model with fixed parameters as a restricted version of the semiparametric model with the unrestricted semiparametric model would indicate whether there is a need for the semiparametric structure. Further, comparison of the DNS-SVt model with the Bayesian semiparametric model would unveil whether the changes in the parameters are in discrete or continuous nature for the typical macro-financial dataset monthly of US bond yields. Following this aim, in this section we discuss several Bayesian model comparison techniques.

Since the key task in our methodology is the density estimation we use the predictive likelihood and Bayes factors computed using the predictive likelihoods 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 previous observations, $y_{1:t_0}$, is given by

$$f(y_{t_0+1}|y_{1:t_0}) = \int p(y_{t_0+1}|f_{t_0+1}, \theta)p(f_{t_0+1}, \theta|y_{1:t_0})df_{t_0+1}d\theta. \quad (18)$$

Notice that we suppress the remaining model parameters for the sake of visibility of the demonstration. The one-step ahead predictive likelihood 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 $t_0 + 1$, can be computed as

$$f(y_{t_0+1}|y_{1:t_0}) \approx \frac{1}{M} \sum_{m=1}^M p(y_{t_0+1}|f_{t_0+1}^m, \theta_{1:t_0}^m), \quad (19)$$

where $p(y_{t_0+1}|f_{t_0+1}^m, \theta_{1:t_0}^m)$ is a multivariate normal density implied in (10) and M is a sufficiently large number. A feature of the predictive likelihoods is that these can be

used to compute the joint predictive likelihood as

$$p(y_{1:t_0}) = \prod_{t=1}^T f(y_{t+1}|y_{1:t}). \quad (20)$$

The joint predictive likelihood can also be interpreted as a measure of marginal likelihood or data density as the distributions of parameters are integrated out. One crucial distinction in computation is that the integration is with respect to posterior distributions rather than prior distributions. Hence, it provides a measure of marginal likelihood that is robust to prior specifications. This provides a tool to analyze the contribution of each observation at time period t to the (log) marginal likelihoods as in (20), see Geweke and Amisano (2010). Bayes factors, then, can be computed as the difference between the marginal likelihoods of the two competing models.

The second metric that we employ is the Root Mean Squared Forecast Error (RMSFE). For forecast horizon h , the RMSFE can be computed as

$$\text{RMSFE}(h) = \sqrt{\left(\frac{1}{M} \sum_{m=1}^M \hat{y}_{t_0+h}^m - y_{t_0+h}\right)^2}, \quad (21)$$

where $\hat{y}_{t_0+h}^m$ is the h -step ahead forecast using the parameter values at step m of the posterior sampler. This metric is conventionally used in many applications to evaluate the forecasting performance of the DNS model, see for example Diebold and Li (2006). Since RMSFE is a tool for measuring the performance of the models in point forecasting rather than density forecast. Together with the Bayes factors we can evaluate the ability of the models both in point and density prediction.

4 Empirical results using US yield data

In this section we examine the potential of the semiparametric framework using a bond yields dataset. Following this we investigate the US treasury bond data. This dataset is analyzed extensively in many applications of the dynamic Nelson-Siegel model, and thus, it provides ample opportunities to compare our 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.¹¹ 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 $T = 480$. We restrict the off-diagonal elements of the autoregressive coefficients matrix to be zero in the main analysis using the DNS framework to keep the model parsimonious,, see Hautsch and Yang (2010) for similar restrictions.¹²

Specifying the decay parameter λ is necessary to complete the DNS model specification. In many applications of the Nelson-Siegel model the value of λ 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 λ prior to estimation facilitates the estimation of the model without losing much explanatory power.

¹¹The 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. We thank Michel van der Wel for making the data available to us.

¹²Marginal likelihoods with mildly informative priors on the off-diagonal elements indicate a high probability mass of these coefficients on the neighborhood of zero.

4.1 Estimation results

The estimates of the factors using the DNS-iMM model are displayed in Figure 1. As yield curve factors are precisely identified for all competing models, the factor estimates are quite similar and we only display those for the iMM 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. Note that the slope of the yield curve, or in other words, term spread, is one of the major leading indicators of the business cycle, see Estrella and Mishkin (1998) for details.

[Insert Figure 1 about here]

We display the estimation results of the DNS-N, DNS-SVt and DNS-iMM models in Table 1 for parameters that are fixed over time with the exception of the variances of the measurement errors which are displayed in Table 2.

[Insert Table 1 about here]

[Insert Table 2 about here]

The results reflect the conventional findings on the US yield curve. First, the persistence of the level factor is considerably high close to a unit root process. The persistence of the remaining factors are lower, though still above 0.9. Interestingly, the persistence of the curvature factor for the model with stochastic volatility and t -distribution is relatively lower. It seems that the explicit modeling of extreme observations and continuously changing volatility of the factors lead to a lower degree of persistence of this factor. Second, the correlation between the factors are quite low for the DNS-N model. Zero is inside the 95% highest posterior density intervals (HPDI) for all correlations confirming the finding that the underlying yield curve

factors are nearly orthogonal to each other, see (Litterman and Scheinkman, 1991) and Diebold *et al.* (2006) for example.

Next, we consider the fixed parameters related to the DNS-SVt model. First, the variances of the time varying intercept are minuscule indicating that there is little variation in the intercept of the yield curve factors. For the volatility process, though, the variation is quite large reflecting the variation of the volatility of the yield curve factors over time. Moreover, the autoregressive parameters are estimated around 0.8 for the volatilities of the slope and curvature factors. Hence, there is a considerable persistence of the volatility processes, albeit from a unit root process, for the slope and curvature factor volatilities.

The estimates of the variances of the measurement errors in Table 2 indicate an increasing precision when the semiparametric structure is used. When the DNS-iMM model is compared to DNS-N and DNS-SVt models it is seen that the variances are decreasing significantly at the short end of the yield curve. While for the medium term yields variances are almost identical, there is a marginal increase in variances at the long end with the use of the semiparametric structure rather than stochastic volatility. In general, the results imply 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.

The evolution of the parameters with the iMM structure and comparison with the estimates for the DNS-N and DNS-SVt models are displayed in Figures 2 and 3.

[Insert Figure 2 about here]

[Insert Figure 3 about here]

The three graphs in Figure 2 correspond to the intercept of the level, slope and curvature factors, respectively. A general finding is that the intercept term is fluctuating around values close to zero irrespective of the model and irrespective of the

fact that whether the intercept is fixed or time varying. This is also due the high persistence of the factors yielding low values of intercept terms. Still, an upward (downward) movement of the intercept term of the level and curvature (slope) factors in the DNS-SVt and DNS-iMM models can be traced, which corresponds to the time periods where the yields attained high values. Nevertheless, these fluctuations are limited in magnitude with the only exceptions around 1980 which are the extreme peaks of the historical interest rates observed during the change of the Federal Reserve's monetary target. It seems that the DNS model with iMM structure categorizes these observations as a separate cluster leading to the interpretation of these observations as outliers. Towards the end of the sample period, the intercepts estimates of the slope and curvature factors (level) for the DNS-SVt model exhibit a downward (upward) adjustment reflecting low values of the yields at the short end of the yield curve during the recent economic downturn. A similar pattern can be seen for the intercept of the slope factor for the DNS-iMM model, albeit quite limited.

The three graphs in Panel A of Figure 3 correspond to the variance of the yield curve factors. The estimation results related to variance parameters of the DNS-iMM and DNS-SVt models are in line with the stylized facts. 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 financial crisis are nicely captured by the time varying volatility regardless whether the changes are modeled discretely or continuously. However, there are two crucial findings. First, the speed of the adjustment of the volatility to the changing conditions differs for each model. When we focus on the variance of the level factor during the Fed's monetary policy experimentation period of 1979-1982 for the DNS-SVt model, for example, we see that the volatility starts to increase around 1979 for both of the models but it reached to its peak rapidly for the DNS-iMM model compared to DNS-SVt model. Although the volatility reaches to its peak around 1981 for the DNS-SVt

model, the level of this peak is much greater compared to DNS-iMM model. Finally, the fade of the volatility lasts until mid 1980's for the DNS-SVt model whereas for the DNS-iMM model the volatility attains low levels already towards the end of 1982 indicating the importance of modeling the volatility with discrete changes for these periods. Secondly, similar to the estimates of the intercept, the estimate of the volatility using DNS-iMM for some specific periods around 1980 is huge again leading to the interpretation of these observations as outliers. Essentially, these periods correspond to March 1980, where there was a jump in the short rate and a sudden but more mediocre increase in the long rate, and to August-September 1981, where there was a large and sudden increase in both long and short rate. For the DNS-SVt model this pattern can also be observed albeit to a much less extent. As a result of the continuously changing volatility there is a mild increase in volatility in March 1980 and the increase continues until September 1981 where volatility reaches to its highest level. Finally, 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 other competing models during the periods of great moderation and with a lower volatility during turbulent times.

The volatility estimates of the slope and curvature factor exhibit a very similar pattern closely resembling to that of the level factor. The recession periods of 1970s, 1980s together with the periods of the Federal Reserve's monetary policy experimentation and the recent recession of 2007-9 are accompanied by high volatility. The volatility remains at low levels for the periods of great moderation as indicated by both DNS-iMM and DNS-SVt models. For the slope factor, there is a jump or an outlier in volatility in March 1980 reflecting the sudden and large increase in the short rate. This increase is mildly captured by mediocre increase in slope factor volatility for the DNS-SVt model. An important finding is on the differences between volatility levels for the DNS-iMM and DNS-SVt models. The volatility estimates

of the slope and curvature factors during tranquil periods are greater for DNS-iMM model than those for the DNS-SVt model. It seems that higher volatility values of the level factor is offset by the lower volatility values of the slope and curvature factors for the DNS-SVt model compared to DNS-iMM model. As in the volatility of the level factor, the volatility estimates of slope and curvature factors for the DNS-N model is close to sample average of the volatility estimates indicated by the DNS-iMM and DNS-SVt models.

When we focus on the estimation results of the correlation coefficients we see that while the estimates of the DNS-N model are close to zero, the estimates for the DNS-iMM model are around 0.40 (0.30) for the correlation between the level and slope (curvature) factors' error terms. An interesting finding is that these correlations decrease significantly or even changing sign during volatile periods. We do not display the 95 % HPDI in Panel B of the Figure 3 to increase the visibility but unreported results indicate that for the correlation estimates the 95% HPDI is very wide with a lower bound very close to zero during calm periods. Moreover, during volatile zero is always inside the 95% HPDI. This indicates that during 'normal' times there is a small but nonzero correlation between level factor and remaining factors but this low correlation breaks down completely during volatility periods. The correlation coefficient between slope and curvature factors are closer to zero with limited fluctuations. Unreported results indicate that zero is always inside the 95% HPDI for both turbulent and calm periods. As in the volatility case, the estimate of the DNS-N model balances between the changing correlations leading to estimates of the correlations close to zero.

Finally, we display the histogram of the number of clusters using the sample from posterior sampler for the DNS-iMM model in Figure 4. The average number of clusters is 3.91 with a median estimate of 4 and mode estimate of 3 as it can be seen from the histogram. Unreported results show that the mode of the distribution

of the number of clusters with more than 60 observations (5 years in total) is 2 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 sampler the recent 2007-9 crisis is identified as a separate cluster distinct from the period before mid 1980s indicating a new type of recession (cluster) though more information is required for more conclusive results, see Çakmaklı *et al.* (2013) for a similar result. This also confirms the findings sofar that the DNS-iMM model identifies two major regimes and the recent recession together with outliers around the beginning of 1980s.

4.2 Predictive performance

The DNS model is often used for its superior predictive performance thanks to its parsimonious yet flexible structure. Therefore, it is crucial to examine the predictive performance of the competing models and the added value of modeling the density of the yield curve factors using discretely or continuously changing parameters compared to the model with fixed parameters. Accordingly, we display the Bayes factors (computed as the difference between the predictive log-likelihood values and that of the DNS-N model using 1-period ahead forecasts) for evaluation of the predictive ability of the models in density forecasting, and the relative RMSFEs (computed as the ratio of the RMSFE of the competing models relative to that of the DNS-N model for 1,3,6 and 12-period forecast horizons) for evaluation of the predictive ability of the models in point forecasting in Table 3.

[Insert Table 3 about here]

We display the Bayes factors for some selected maturities in the first three columns of Table 3. There are two clear-cut conclusions. The Bayes factors relative to the DNS-N model is highest and it is of magnitude of around 150-200 points for the DNS-iMM model regardless of the duration of maturities. This implies that the

DNS-iMM model possess much superior density forecasting ability than the DNS-N model. DNS-SVt also attains a large and positive value of Bayes factor relative to the DNS-N model for all maturities, nevertheless, the magnitude of the Bayes factors is of 50-100 and thus the predictive ability of the DNS-SVt model is considerably lower than that of the DNS-iMM model for all maturities. Therefore, for all maturities, the predictive likelihood values monotonically increase starting from DNS-N model to DNS-SVt and then to DNS-iMM models. Second, the Bayes factor values decrease almost monotonically with the increase in the duration of the maturity of the yields. Hence, it is relatively more difficult to beat the DNS-N model at the long end of the yield curve regardless of the type of the models. When we focus on the short end of the yield curve we see a substantial improvement of the predictive likelihoods for DNS-SVt model compared to DNS-N model as it can be seen from the large and positive Bayes factor values. This shows the importance time variation in the parameters of the yield curve factors. In addition when these changes are modeled discretely using the DNS-iMM model, the improvement in the predictive likelihoods is substantial. For 3 months maturity, for example, the Bayes factor for the DNS-SVt model is 71.2. However, this value jumps to 197.4 when we consider the Bayes factor for DNS-iMM model. For the medium term maturities this picture prevails. When we consider 24 month maturity, for example, the Bayes factor for the DNS-SVt model becomes 73.5, whereas the Bayes factor for the DNS-iMM model is as high as 167.5. A similar pattern can also be observed for the long term maturities. The Bayes factor for 120 months maturities is 32.4 when we consider the DNS-SVt model and it increases substantially to 130.5 for the DNS-iMM model.

We display the evolution of the joint predictive log-likelihood values at each period as the sum of the individual likelihoods over time starting from January 1982 for the selected maturities 3, 12, 24, 60, 120 months, representing short, medium and long term maturities in Figure 5. Here we display predictive log-likelihood

values rather than Bayes factors to provide extra information on the predictive log-likelihood values, nevertheless the discrepancies between the lines over time reflects the evolution of the Bayes factors over time.

[Insert Figure 5 about here]

Figure 5 reveals two important findings. First, the speed of adjustment to the new conditions is much faster and accurate for the DNS-iMM model compared to DNS-SVt and DNS-N models. While this finding is common to all maturities, it is most pronounced for longer maturities. For example, the downturn of the predictive likelihoods during 2008, that is, the deterioration of the predictive abilities of the models at the onset of the recession, is mildest and shortest lasting for the DNS-iMM model compared to remaining models for the 120 months maturity as the bottom graph of Figure 5 shows. Indeed, predictive likelihood values only for the period starting from January 2006 until December 2009, for 120 months maturity, become -15.7, -13.4 and -2.8 for the DNS-N, DNS-SVt and DNS-iMM models, respectively. When we perform the same computation for the 3 months maturities, the predictive likelihood values becomes -46.9, -42.9 and -23.4 for the DNS-N, DNS-SVt and DNS-iMM models, respectively.

Second, the predictive likelihood values during calm periods are much higher for the DNS-iMM model compared to DNS-SVt model and DNS-N models. While, all five graphs share this common feature, it is even more pronounced for the long maturities. The discrepancy between the predictive likelihoods of the DNS-iMM models and those of the DNS-SVt and DNS-N models increases constantly during the periods following the turbulent times of FED's monetary policy experiments at the beginning of 1980s as well as the recent economic downturn. Combined with the first finding on the superior speed of adjustment of the DNS-iMM model to the changing conditions, this comparison between the models indicate the fact that when some parts of yields data exhibit unconventional changes it may cause to ill-behaved

regions in the posterior distributions blurring the econometric inference. Semiparametric structure in the DNS framework, i.e. DNS-iMM model nicely captures this part as outliers or as separate clusters with distinct dynamics, leading to the inference about the remaining parts to be unaffected. This tendency can be observed clearly for all maturities in the sense that the discrepancy between the predictive likelihood values, i.e. Bayes factor, is becoming wider in favour of the the DNS-iMM model over the course of the calm periods indicating greater contribution of the individual likelihoods to the joint likelihood for the post 1980 periods also called as great moderation.

Finally, we display the RMSFE's of models relative to the RMSFE of the DNS-N model for each maturity in Table 3 for 1, 3, 6 and 12 months forecast horizons to evaluate the ability of the models in point forecasting. First, it is seen that for 1 and 3 months forecast horizon, the DNS-SVt and DNS-iMM models perform very close to each other and both models have a much superior performance compared to DNS-N model. The RMSFE of these models is almost half of the RMSFE of the DNS-N model for the 1 month forecast horizon and almost three fourth of the RMSFE of the DNS-N model for the 3 months horizon. While at the shortest and longest end of the yield curve, this relative predictive power deteriorate slightly, still these models have much higher predictive ability compared to DNS-N model for these horizons. It seems that the additional flexibility brought by the time varying parameters of the DNS-SVt and DNS-iMM models leads to better predictive power compared to more parsimonious DNS-N model in the short horizons. However, this picture reverses when we consider 6 and 12 months horizon. The DNS-SVt model has just slightly better performance compared to DNS-N model well above 0.9 relative RMSFE for all maturities except the 120 month maturity for which the DNS-N model outperforms DNS-SVt model. For the 12 months forecast horizon, actually DNS-N model performs much better than DNS-SVt model for all horizons. Hence, the

additional flexibility of the DNS-SVt model is penalized at long horizons compared to more parsimonious DNS-N model. A striking finding is that this pattern is not the case for the DNS-iMM model. Although the relative predictive performance worsens slightly compared to shorter horizons, it still outperforms the DNS-N model for 6 and 12 months as well. This shows that modelling the time variation in the parameters of yield curve factors in a discrete rather than continuous manner provides additional flexibility by still preserving sufficient level of parsimony as it can be seen from the relative RMSFE's at both short and longer forecast horizons. Such a balance is not the case for the DNS-SVt model, since the superior predictive power erodes rapidly with the increasing forecast horizons relative to parsimonious models as DNS-N model.

An obvious reason of the outperformance of the models with time varying parameters is the periods with distinct characteristics in our estimation sample. Since the switch from turbulent times of 1970s relative to great moderation period requires the parameters of the competing models to alter as well, models with time varying parameters can perform better in forecasting especially at shorter horizons. To examine the forecasting performance of the models during the great moderation period, we repeat the forecasting exercise using the sample starting from January 1985. The results are displayed in Table 4.

[Insert Table 4 about here]

When we consider density prediction in the first three columns of Table 4, we see that both DNS-SVt and DNS-iMM model still do much better than the DNS-SVt model as positive Bayes factors indicate.¹³ This indicates that modelling density explicitly indeed pays off. As in the full sample results DNS-iMM model outperforms DNS-SVt model reflecting the importance of capturing discrete changes in the parameters

¹³Notice that the diminishing magnitude of the Bayes factors is mainly due to reduced sample size.

also in the post-1985 period. Hence, superior density forecasting performance of the DNS-iMM model is robust and independent of the sample period chosen.

When we consider point prediction results with the changing sample period, we observe that the performance of the DNS-SVt and DNS-iMM model deteriorates. For 1-period ahead forecasts, all three competing models perform very close to each other with the DNS-iMM performing best with a marginal difference for the short and long end of the yield curve. At the medium term, DNS-SVt model outperforms but only marginally. As in the previous case the performance of the DNS-SVt model worsens substantially with the increasing forecast horizon. For the yields with 60 months maturity, for example, the relative RMSFE of 0.99 for the 1-period ahead forecasts reduces to only 1.51 for the 12-period ahead forecasts. Conversely, an astonishing finding is on the performance of the DNS-iMM model in the sense that the relative RMSFEs actually improves with the increasing forecast horizon. For the yields with 60 months maturity, the relative RMSFE of 0.96 for the 1-period ahead forecasts increases to 0.75 for the 12-period ahead forecasts. This indicates that the additional flexibility of the DNS-iMM model is not at a cost of reducing parsimony. Overall, the findings related to DNS-iMM are quite robust and do not change with the exclusion of the periods of turbulent times.

5 Conclusion

In this paper, we put forward the Bayesian semiparametric dynamic Nelson-Siegel (DNS-iMM) model to estimate the density of the yield curve factors using the dynamic Nelson-Siegel model. Comparison of the model with the more conventional versions of dynamic Nelson-Siegel model with fixed parameters (DNS-N) as well as with continuously changing intercept and volatility parameters coupled with t distribution of the yield curve factors (DNS-SVt) using US monthly zero coupon bond yields over the period from January 1970 until December 2009 reveals many

important findings.

The parameters of the underlying yield curve factors vary over time to accommodate the changing economic conditions as documented also in other studies. This variation is specifically important for the second moment of the factors as the estimation results of the DSN-SVt and DNS-iMM indicate. However, modelling this variation in a discrete manner as in the DNS-iMM model rather than in a continuous manner is crucial. First, the speed of adjustment of the varying parameters to the rapidly changing conditions is much superior and accurate for the DNS-iMM model compared to DNS-SVt model. This is reflected as superior Bayes factors for density forecasts and superior RMSFEs relative to DNS-N models during turbulent times such as the periods of Federal Reserves monetary policy experimentation of 1979-82 and the beginning of 1980s as well as the recent 2007-9 financial crisis. Second, the number of clusters with distinct parameters in the DNS-iMM model is estimated as four, implying that some of the clusters capture aberrant observations or outliers efficiently compared to the stochastic volatility and the t distribution in the DNS-SVt model. Modelling the dynamics during the turbulent times and aberrant observations accurately also enables us to represent more tranquil periods with much precise parameter values. This is reflected with much higher Bayes factors and relative RMSFE values during calm periods for the DNS-iMM model. An important result of this finding is that the DNS-iMM model preserves a desirable balance between additional flexibility and parsimony. This is not the case for the DNS-SVt model as the superior forecasting performance of the DNS-SVt model is only the case for the short horizons but it deteriorates rapidly for the longer horizons. However, the DNS-iMM model outperforms DNS-SVt model as well as parsimonious DNS-N model for all forecast horizons indicating the importance of this balance.

We also perform a similar forecasting exercise using the sample starting from 1985, i.e. excluding the pre-'great moderation' period. While the DNS-iMM model

still outperforms competing models in density forecasting, relative RMSFEs show a similar performance of the competing models at short horizons. Still at longer forecast horizons, DNS-iMM model provides superior point forecasting performance. This shows that the semiparametric structure of the DNS-iMM model possess a flexible yet parsimonious structure that is robust to various kinds of datasets.

Since the model falls into the class of dynamic factor models, which in turn is a special case of general state-space models, the methodology in this paper provides a unified approach to model different types of non-Gaussian behaviour and discrete changes in model parameters in a wide variety of applications. Therefore, the model and the resulting estimation methodology proposed in this paper can be extended in many directions. An interesting next step is inclusion of the macroeconomic variables such as inflation and industrial production in the VAR system of state equations. Such an approach could reveal the nonlinear and rapidly changing relation between the yield curve factors and these variables with the changing conditions.

References

- Antoniak, C. E. (1974), Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems, *The Annals of Statistics*, 2, 1152–1174.
- Bianchi, F., H. Mumtaz, and P. Surico (2009), Dynamics of the term structure of UK interest rates, Bank of England working papers 363, Bank of England.
- Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307–327.
- Çakmaklı, C., R. Paap, and D. van Dijk (2013), Measuring and predicting heterogeneous recessions, *Journal of Economic Dynamics and Control*, 37, 2195–2216.
- Carriero, A., T. Clark, and M. Marcellino (2014), No Arbitrage Priors, Drifting Volatilities, and the Term Structure of Interest Rates, CEPR Discussion Papers 9848, C.E.P.R. Discussion Papers.
- Carter, C. K. and R. Kohn (1994), On Gibbs sampling for state space models, *Biometrika*, 81, 541–553.
- Chib, S. and B. H. Hamilton (2002), Semiparametric Bayes analysis of longitudinal data treatment models, *Journal of Econometrics*, 110, 67–89.
- Christensen, J. H., F. X. Diebold, and G. D. Rudebusch (2011), The affine arbitrage-free class of Nelson-Siegel term structure models, *Journal of Econometrics*, 164, 4–20.
- 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, 276–305.
- Cox, J. C., J. E. J. Ingersoll, and S. A. Ross (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*, 53, 385–407.
- Creal, D. (2012), A survey of sequential Monte Carlo methods for economics and finance, *Econometric reviews*, 31, 245–296.
- Curdia, V., M. del Negro, and D. L. Greenwald (2014), RARE SHOCKS, GREAT RECESSIONS, *Journal of Applied Econometrics*, 29, 1031–1052.
- Dai, Q. and K. J. Singleton (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance*, 55, 1943–1978.
- Dai, Q. and K. J. Singleton (2002), Expectation puzzles, time-varying risk premia, and affine models of the term structure, *Journal of Financial Economics*, 63, 415–441.
- De Pooter, M. (2007), Examining the Nelson-Siegel Class of Term Structure Models, Tinbergen Institute Discussion Papers 07-043/4, Tinbergen Institute.

- Diebold, F. X. and C. Li (2006), Forecasting the term structure of government bond yields, *Journal of Econometrics*, 130, 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, 309–338.
- Duffee, G. R. (2002), Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance*, 57, 405–443.
- Duffie, D. and R. Kan (1996), A Yield-Factor Model Of Interest Rates, *Mathematical Finance*, 6, 379–406.
- Durbin, J. and S. Koopman (2012), *Time Series Analysis by State Space Methods: Second Edition*, Oxford Statistical Science Series, OUP Oxford.
- Estrella, A. and F. S. Mishkin (1998), Predicting U.S. Recessions: Financial Variables As Leading Indicators, *The Review of Economics and Statistics*, 80, 45–61.
- Fama, E. F. and R. R. Bliss (1987), The Information in Long-Maturity Forward Rates, *American Economic Review*, 77, 680–92.
- Friedman, M. (1984), Lessons from the 1979-82 Monetary Policy Experiment, *American Economic Review*, 74, 397–400.
- Frühwirth-Schnatter, S. (1994), Data augmentation and dynamic linear models, *Journal of Time Series Analysis*, 15, 183–202.
- Geisser, S. (1965), A Bayes Approach for Combining Correlated Estimates, *Journal of the American Statistical Association*, 602–607.
- 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. (1993), Bayesian Treatment of the Independent Student- t Linear Model, *Journal of Applied Econometrics*, 8, 19–40.
- Geweke, J. (2005), *Contemporary Bayesian econometrics and statistics*, Wiley series in probability and statistics, John Wiley.
- Geweke, J. and G. Amisano (2010), Comparing and evaluating Bayesian predictive distributions of asset returns, *International Journal of Forecasting*, 26, 216–230.
- Griffin, J. E. and M. F. J. Steel (2004), Semiparametric Bayesian inference for stochastic frontier models, *Journal of Econometrics*, 123, 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, 2291–2304.

- Hautsch, N. and Y. Ou (2009), Analyzing Interest Rate Risk: Stochastic Volatility in the Term Structure of Government Bond Yields, CFS Working Paper Series 2009/03, Center for Financial Studies.
- Hautsch, N. and F. Yang (2010), Bayesian inference in a Stochastic Volatility Nelson-Siegel model, *Computational Statistics & Data Analysis*, In Press, Corrected Proof, –.
- Hevia, C., M. Gonzalez-Rozada, M. Sola, and F. Spagnolo (2015), Estimating and Forecasting the Yield Curve Using A Markov Switching Dynamic Nelson and Siegel Model, *Journal of Applied Econometrics*, 30, 987–1009.
- Hirano, K. (2002), Semiparametric Bayesian Inference in Autoregressive Panel Data Models, *Econometrica*, 70, 781–799.
- Hull, J. and A. White (1990), Pricing Interest-Rate-Derivative Securities, *Review of Financial Studies*, 3, 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, 185–212.
- Jensen, M. J. (2004), Semiparametric Bayesian Inference of Long-Memory Stochastic Volatility Models, *Journal of Time Series Analysis*, 25, 895–922.
- Jensen, M. J. and J. M. Maheu (2010), Bayesian semiparametric stochastic volatility modeling, *Journal of Econometrics*, 157, 306–316.
- Kim, C.-J. and C. R. Nelson (1999), *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, vol. 1, 1 edn., The MIT Press.
- 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, 329–343.
- Litterman, R. B. and J. Scheinkman (1991), Common Factors Affecting Bond Returns, *The Journal of Fixed Income*, 1, 54–61.
- McCulloch, J. H. (1975), The Tax-Adjusted Yield Curve, *Journal of Finance*, 30, 811–30.
- Mesters, G., B. Schwaab, and S. J. Koopman (2014), A Dynamic Yield Curve Model with Stochastic Volatility and Non-Gaussian Interactions: An Empirical Study of Non-standard Monetary Policy in the Euro Area, Tinbergen Institute Discussion Papers 14-071/III, Tinbergen Institute.
- Neal, R. M. (2003), Slice sampling, *The Annals of Statistics*, 31, 705–767.

- Nelson, C. R. and A. F. Siegel (1987), Parsimonious Modeling of Yield Curves, *Journal of Business*, 60, 473–89.
- Nelson, C. R. and A. F. Siegel (1988), Long-Term Behavior of Yield Curves, *Journal of Financial and Quantitative Analysis*, 23, 105–110.
- Nyholm, K. and R. Rebonato (2008), Long-horizon yield curve projections: comparison of semi-parametric and parametric approaches, *Applied Financial Economics*, 18, 1597–1611.
- Oh, D. H. and A. J. Patton (2016), Modelling Dependence in High Dimensions with Factor Copulas, *Journal of Business & Economic Statistics*, 'forthcoming'.
- Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007), Stochastic volatility with leverage: Fast and efficient likelihood inference, *Journal of Econometrics*, 140, 425–449.
- Orphanides, A. and D. H. Kim (2005), Term Structure Estimation with Survey Data on Interest Rate Forecasts, Computing in Economics and Finance 2005 474, Society for Computational Economics.
- Primiceri, G. E. (2005), Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies*, 72, 821–852.
- Sethuraman, J. (1994), A constructive definition of Dirichlet priors, *Statistica Sinica*, 4, 639–650.
- Stock, J. H. and M. W. Watson (2007), Why Has U.S. Inflation Become Harder to Forecast?, *Journal of Money, Credit and Banking*, 39, 3–33.
- Svensson, L. E. (1994), Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994, NBER Working Papers 4871, 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.
- Van Gael, J., Y. Saatchi, Y. W. Teh, and Z. Ghahramani (2008), Beam Sampling for the Infinite Hidden Markov Model, in *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, ACM, New York, NY, USA.
- Vasicek, O. (1977), An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177–188.
- West, M. (1992), Hyperparameter estimation in Dirichlet process mixture models, ISDS Discussion Paper 92-A03, Duke University.
- Xiang, J. and X. Zhu (2013), A Regime-Switching Nelson-Siegel Term Structure Model and Interest Rate Forecasts, *Journal of Financial Econometrics*.

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, 579–591.

Tables and Figures

Table 1: Posterior results for competing models

	DNS-N			DNS-SVt			DNS-iMM		
	Level	Slope	Curvature	Level	Slope	Curvature	Level	Slope	Curvature
μ_f	0.123 (0.045)	-0.080 (0.038)	-0.068 (0.044)						
ϕ_f	0.983 (0.005)	0.956 (0.016)	0.911 (0.021)	1	0.948 (0.019)	0.843 (0.031)	0.990 (0.000)	0.976 (0.009)	0.915 (0.018)
σ_f^2	0.221 (0.015)	0.431 (0.029)	0.830 (0.065)						
ρ_{1j}	1	0.074 (0.047)	-0.012 (0.049)						
ρ_{2j}	0.074 (0.047)	1	0.049 (0.050)						
ρ_{3j}	-0.012 (0.049)	0.049 (0.050)	1						
$\sigma_{\mu_i}^2$				0.001 (0.000)	0.001 (0.000)	0.001 (0.000)			
ϕ_{h_i}				1	0.788 (0.207)	0.791 (0.173)			
$\sigma_{h_i}^2$				0.243 (0.010)	0.207 (0.060)	0.215 (0.062)			
# of clusters								3.91 (1.44)	

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-SVt is the model with innovations following a t distribution with a stochastic volatility. DNS-iMM is the model with Infinite Markov Mixture structure, where both the mean and the variance of the innovations have a hierarchical DP prior. μ_f for $f = 1, 2, 3$ is the unconditional mean of the factor errors (intercept of the factors) and ϕ_f is the autoregressive coefficients of the factors. σ_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 $j \in \{1, 2, 3, \setminus i\}$. ϕ_{h_i} is the autoregressive coefficients of the factor volatilities, $\sigma_{\mu_i}^2$ is the variance of the time varying factor intercepts and $\sigma_{h_i}^2$ is the variance of the factor volatilities for the DNS-SVt model. The values indicate the posterior means of the parameters with the posterior standard deviations in parenthesis. Marginal likelihood values are computed using (20). 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 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 for selected maturities

	<u>DNS-N</u>	<u>DNS-SVt</u>	<u>DNS-iMM</u>
M3	12.74 (0.96)	12.72 (0.96)	12.38 (0.92)
M6	2.22 (0.22)	2.22 (0.21)	2.08 (0.20)
M9	0.81 (0.09)	0.80 (0.09)	0.77 (0.08)
M12	0.54 (0.05)	0.53 (0.05)	0.56 (0.05)
M15	0.72 (0.06)	0.72 (0.06)	0.76 (0.06)
M18	0.65 (0.05)	0.66 (0.05)	0.68 (0.05)
M21	0.63 (0.05)	0.63 (0.05)	0.65 (0.05)
M24	0.41 (0.03)	0.42 (0.03)	0.42 (0.03)
M36	0.20 (0.02)	0.19 (0.02)	0.20 (0.02)
M48	0.35 (0.03)	0.35 (0.03)	0.36 (0.03)
M60	0.41 (0.03)	0.41 (0.03)	0.41 (0.03)
M120	4.48 (0.34)	4.47 (0.34)	4.50 (0.34)

Note: Mx indicates the yields with maturity x months. The results show the posterior means and posterior standard deviations (in parenthesis) of the standard deviations of the measurement errors (multiplied by hundred) in (2), i.e. the diagonal elements of R . See Table 1 for further details.

Table 3: Bayes factors and relative RMSFE's of the competing models for selected maturities

	<u>Bayes factor</u>			<u>Relative RMSFE's</u>											
	N	SVt	iMM	<u>$h = 1$</u>			<u>$h = 3$</u>			<u>$h = 6$</u>			<u>$h = 12$</u>		
				N	SVt	iMM	N	SVt	iMM	N	SVt	iMM	N	SVt	iMM
M3	0	71.2	197.4	1	0.58	0.54	1	0.73	0.69	1	0.97	0.81	1	1.20	0.88
M6	0	102.1	214.9	1	0.48	0.46	1	0.69	0.66	1	0.94	0.80	1	1.19	0.88
M9	0	97.9	203.0	1	0.46	0.46	1	0.68	0.65	1	0.93	0.79	1	1.18	0.87
M12	0	91.8	189.3	1	0.45	0.46	1	0.68	0.66	1	0.92	0.80	1	1.17	0.87
M15	0	85.7	180.1	1	0.45	0.48	1	0.68	0.66	1	0.92	0.80	1	1.17	0.88
M18	0	81.7	174.6	1	0.45	0.48	1	0.68	0.66	1	0.91	0.80	1	1.17	0.87
M21	0	78.0	169.4	1	0.46	0.48	1	0.68	0.67	1	0.91	0.80	1	1.17	0.87
M24	0	73.5	167.5	1	0.47	0.49	1	0.69	0.67	1	0.92	0.80	1	1.17	0.86
M36	0	68.4	162.5	1	0.49	0.50	1	0.71	0.68	1	0.92	0.79	1	1.17	0.86
M48	0	55.4	153.7	1	0.52	0.53	1	0.74	0.70	1	0.95	0.81	1	1.20	0.86
M60	0	45.6	150.3	1	0.55	0.55	1	0.76	0.71	1	0.97	0.81	1	1.22	0.87
M120	0	32.4	130.5	1	0.72	0.71	1	0.88	0.81	1	1.07	0.89	1	1.31	0.94

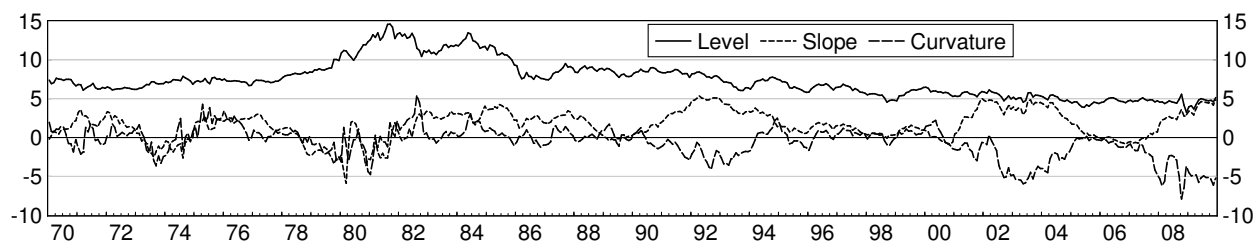
Note: Mx indicates the yields with maturity of x months. 'N' stands for the model where the distribution of the factors are assumed to be Gaussian with fixed parameters. 'SVt' stands for the model with stochastic volatility for the volatility of the factors coupled with a t -distribution. 'iMM' stands for the model with Infinite Markov Mixture structure for the conditional distribution of the factors. See Table 1 for further details. RMSFE is short for root mean square forecast error and h represents the horizon of the forecast. The evaluation sample spans the period over 1982-January until 2009-December.

Table 4: Bayes factors and relative RMSFE's of the competing models for selected maturities using the post 1985 sample

	Bayes factor			Relative RMSFE's											
	N	SVt	iMM	$h = 1$			$h = 3$			$h = 6$			$h = 12$		
				N	SVt	iMM	N	SVt	iMM	N	SVt	iMM	N	SVt	iMM
M3	0	24.72	90.78	1.00	1.01	0.93	1.00	0.98	0.93	1.00	1.22	0.92	1.00	1.42	0.92
M6	0	35.76	82.63	1.00	0.99	0.92	1.00	0.96	0.95	1.00	1.29	0.93	1.00	1.48	0.92
M9	0	27.93	71.87	1.00	0.98	0.94	1.00	0.97	0.96	1.00	1.34	0.93	1.00	1.51	0.92
M12	0	21.18	63.73	1.00	0.96	0.97	1.00	0.96	0.98	1.00	1.35	0.93	1.00	1.51	0.91
M15	0	7.56	65.60	1.00	0.95	0.98	1.00	0.96	0.98	1.00	1.34	0.92	1.00	1.51	0.89
M18	0	14.72	63.04	1.00	0.96	0.98	1.00	0.97	0.98	1.00	1.35	0.91	1.00	1.51	0.88
M21	0	16.42	62.66	1.00	0.97	0.98	1.00	0.98	0.99	1.00	1.36	0.91	1.00	1.51	0.87
M24	0	9.42	61.06	1.00	0.97	0.98	1.00	0.98	0.98	1.00	1.37	0.91	1.00	1.52	0.85
M36	0	13.54	56.80	1.00	0.99	0.98	1.00	1.00	0.98	1.00	1.36	0.88	1.00	1.52	0.81
M48	0	20.94	84.29	1.00	1.00	0.98	1.00	1.01	0.99	1.00	1.38	0.88	1.00	1.56	0.78
M60	0	-0.94	66.21	1.00	0.99	0.96	1.00	1.02	0.99	1.00	1.43	0.87	1.00	1.65	0.75
M120	0	-2.52	67.93	1.00	1.03	1.01	1.00	1.07	1.05	1.00	1.71	0.99	1.00	2.08	0.85

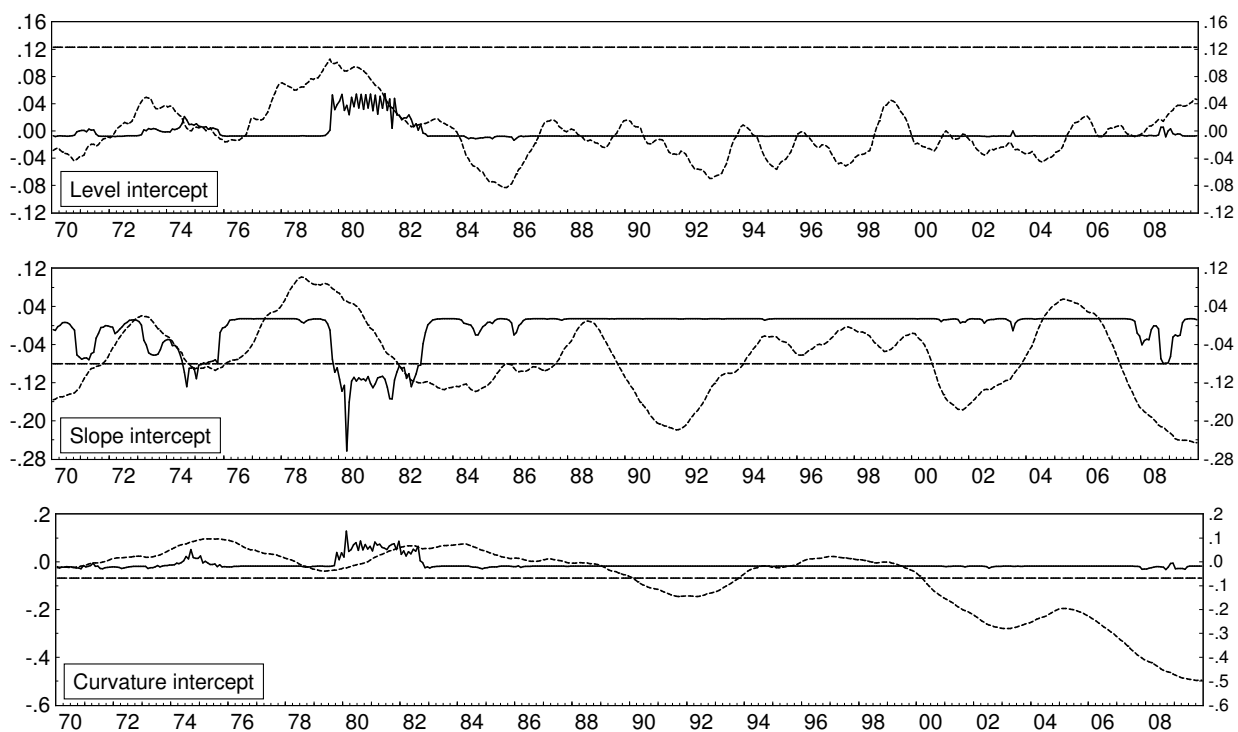
Note: Mx indicates the yields with maturity of x months. 'N' stands for the model where the distribution of the factors are assumed to be Gaussian with fixed parameters. 'SVt' stands for the model with stochastic volatility for the volatility of the factors coupled with a t -distribution. 'iMM' stands for the model with Infinite Markov Mixture structure for the conditional distribution of the factors. See Table 1 for further details. RMSFE is short for root mean square forecast error and h represents the horizon of the forecast. The evaluation sample spans the period over 1993-January until 2009-December.

Figure 1: Posterior means of the factors for the DNS-iMM model



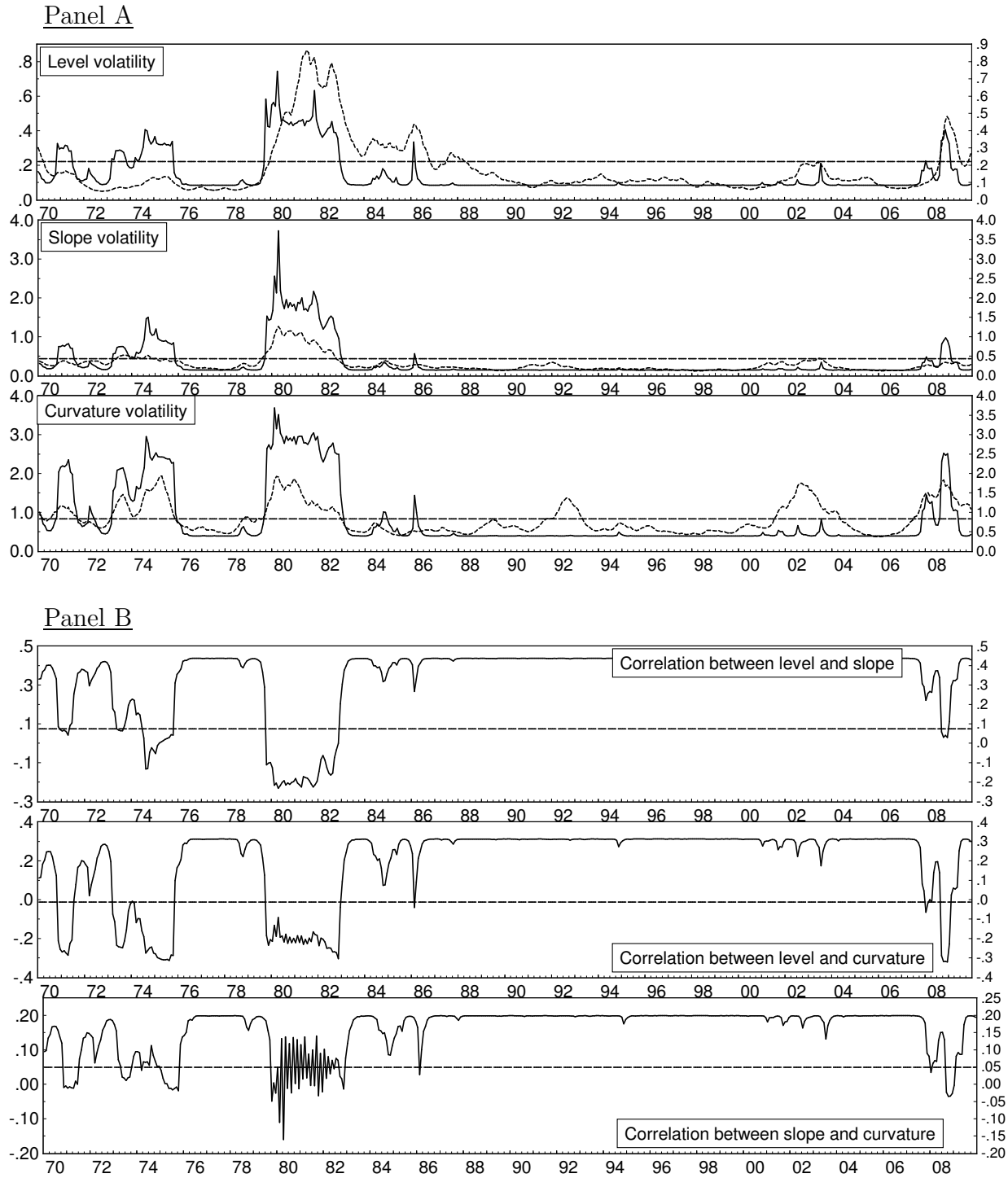
Note: See Table 1 for abbreviations. The solid, the dotted, and the dashed lines represent posterior means of the level, slope, and curvature factors, respectively, estimated using the DNS-iMM model.

Figure 2: Posterior means of the factor intercept parameters for DNS-iMM (solid line), DNS-SVt (dotted line) and DNS-N (dashed line) models



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

Figure 3: Posterior means of the factor variance and correlation parameters for DNS-iMM (solid line), DNS-SVt (dotted line) and DNS-N (dashed line) models



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

Figure 4: The histogram of the number of clusters using 50,000 draws from the posterior sampler

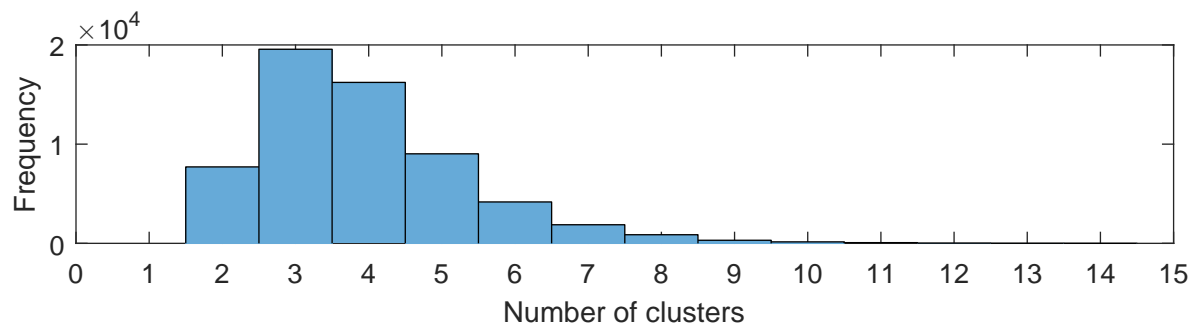
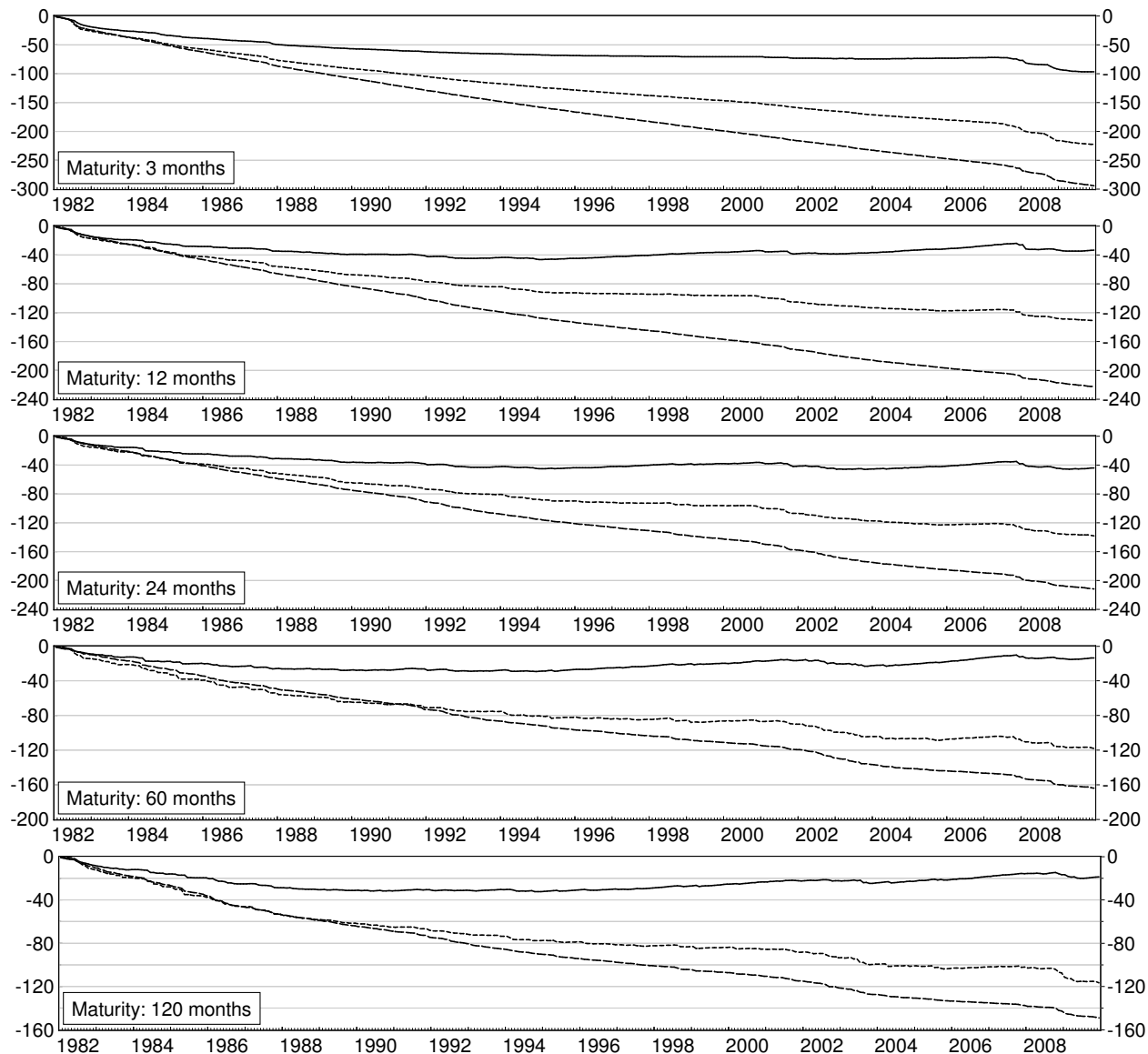


Figure 5: Evolution of the marginal likelihood values of DNS-iMM (solid line), DNS-SVt (dotted line) and DNS-N models (dashed line) for selected maturities over time



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-SVt model and the dashed lines represent the corresponding values for DNS-iMM model. See Table 1 for further details.

Appendices

A Full conditional posterior distributions

In this section we describe the full conditional posterior distributions of the parameters that are common to both the model with semiparametric structure and the model with only 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. In Section A.1 we provide details on the full conditional posterior distributions of the common fixed parameters for the competing models as these are identical conditional on the remaining parameters. The full conditional posterior distributions of the parameters for the DNS-SVt model that are not fixed are discussed in Section A.2.

A.1 Sampling of common fixed parameters for the competing models

A.1.1 Sampling of autoregressive parameters, F

Conditional on the yield curve factors $f_{0:T}$ and the parameters $\theta_{1:T}$ where $\theta_t = \{\mu_t, Q_t\}$ the state equation in (3) can be rewritten as a Vector AutoRegression (VAR) of first order as follows

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

Redefining the variables such that $z_t = f_t - \mu_t$ and $x_t = f_{t-1}$ for $t = 1, \dots, T$, (A.1) can be rewritten as

$$z_t = F x_t + \varepsilon_t \text{ and } \varepsilon_t \sim N(0, Q_t). \quad (\text{A.2})$$

In compact form the model becomes

$$Z = FX + U, \quad (\text{A.3})$$

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

$$\begin{aligned} \text{vec}(Z) &= (X' \otimes I_3) * \text{vec}(F) + \text{vec}(U) \\ z &= (X' \otimes I_3) * \text{vec}(F) + u \end{aligned} \quad (\text{A.4})$$

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

Using the multivariate normal prior $N(\underline{\mu}_{\text{vec}(F)}, \underline{\Sigma}_{\text{vec}(F)})$ in (13) 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

$$\begin{aligned} \bar{\Sigma}_{\text{vec}(F)} &= \left(\underline{\Sigma}_{\text{vec}(F)}^{-1} + (X' \otimes I_3)' \Omega^{-1} (X' \otimes I_3) \right)^{-1} \\ \bar{\mu}_{\text{vec}(F)} &= \bar{\Sigma}_{\text{vec}(F)} \left(\underline{\Sigma}_{\text{vec}(F)}^{-1} \underline{\mu}_{\text{vec}(F)} + (X' \otimes I_3)' \Omega^{-1} (X' \otimes I_3) \text{vec}(F)_{GLS} \right), \end{aligned} \quad (\text{A.5})$$

where $\text{vec}(F)_{GLS} = ((X' \otimes I_3)' \Omega^{-1} (X' \otimes I_3))^{-1} (X' \otimes I_3)' \Omega^{-1} z$. Notice that the parameters $\theta_{1:T}$ are governed by the semiparametric structure for the DNS-iMM model and they follow a pure random walk and an autoregressive processes for the DNS-SVt model. Finally, they are fixed for the DNS-N model.

A.1.2 Sampling of covariance matrix of the yields, R

Given that the yield curve factors are sampled at the m^{th} step of the posterior sampler, and thus, they are observed, the measurement equation can be written as a standard regression as

$$y_t = H f_t + \eta_t. \quad (\text{A.6})$$

Since R is the covariance matrix of η_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 $\bar{S}_R = \underline{S}_R + \sum_{t=1}^T (y_t - Hf_t)(y_t - Hf_t)'$.

A.1.3 Sampling of yield curve factors, f_t

Conditional on the remaining model parameters, drawing $f_{0:T}$ can be implemented using standard inference of the linear Gaussian state-space models. 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). As this has become a standard practice in many applications, here we do not provide a detailed analysis but refer to standard textbooks such as Durbin and Koopman (2012).

A.2 Sampling of parameters related to DNS-SVt model

A.2.1 Sampling of volatilities, $h_{i,t}$

Given that the yield curve factors are sampled at the m^{th} step of the posterior sampler, and thus, they are observed, sampling of the volatilities is identical to the standard case as in Omori *et al.* (2007). Briefly, this involves approximating the $\log\text{-}\chi^2$ distribution due to nonlinear transformation of the model using a mixture of ten Gaussian distributions. Hence, conditional on the mixtures the model becomes linear and Gaussian where standard inference can be carried out easily. Here, we also proceed with the same strategy as in Omori *et al.* (2007) and refer to this study for details. Alternatively, sequential Monte Carlo methods serve as an attractive strategy that provides exact inference. Indeed, Rao-Blackwellized particle filters is very suitable for the problem at hand, see, for example, Creal (2012). However, in our estimation, the focus is mainly on the smoothed estimates of the volatility rather than online estimation. Therefore, we do not perform sequential Monte Carlo methods as

they are computationally costly, especially, when smoothed estimates are required. Given that we perform a recursive forecasting exercise, such computationally costly approaches are not feasible.

A.2.2 Sampling of volatility and intercept variances, $\sigma_{h_i}^2$ and $\sigma_{\mu_i}^2$

Conditional on the volatilities, h_t (intercepts, μ_t) and the state equations in (9), posterior distribution of the variance, σ_{h_i} (σ_{μ_i}) for $i = 1, 2, 3$ reduces to the case of a standard linear model. Hence, using standard results from a linear regression model with an inverse-Gamma prior with scale parameter \underline{S}_h (\underline{S}_μ) and degrees of freedom \underline{T}_h (\underline{T}_μ) for the variance, $\sigma_{h_i}^2$ ($\sigma_{\mu_i}^2$), it follows that the conditional posterior distribution of $\sigma_{h_i}^2$ ($\sigma_{\mu_i}^2$) is an inverse-Gamma distribution with scale parameter $\sum_{t=2}^T \eta_{i,t}^{\mu^2} + \underline{S}_\mu$ ($\sum_{t=2}^T \eta_{i,t}^{h^2} + \underline{S}_h$) and with $T + \underline{T}_h$ ($+\underline{T}_\mu$) degrees of freedom, from which draws can be sampled easily. For the variance of the volatility we impose strong priors around the value of 0.25 where we conduct a prior predictive analysis for the choice of the nformave prior. Note that the choice of the prior is very similar to Stock and Watson (2007) who employ a similar modeling approach using US inflation data. For the variance of the intercept we use non-informative priors as likelihood information dominates prior information according to our experimentation.

A.2.3 Sampling of autoregressive parameters, ϕ_{h_i}

Conditional on the volatilities, ϕ_{h_i} and the state equations in (9), posterior distribution of ϕ_{h_i} for $i = 1, 2, 3$ reduces to the case of a standard linear model. Hence, using standard results from a linear regression model with a Normal prior, it follows that the conditional posterior distribution of ϕ_{h_i} is a Normal distribution with variance $\bar{V}_i = (\sum_{t=2}^T h_{i,t-1}^2 / \sigma_{h_i}^2 + \underline{V}^{-1})^{-1}$ and mean $\bar{m}_i = \bar{V}^{-1}(\underline{V}^{-1}\underline{m} + (\sum_{t=2}^T h_{i,t-1}^2 / \sigma_{h_i}^2)\phi_{h_i,OLS})$, where $\phi_{h_i,OLS}$ is the OLS estimator and \underline{m} and \underline{V} are the prior mean and variance. We can be sample from $N(\bar{m}_i, \bar{V}_i)$ accordingly.

A.2.4 Modeling t - distribution

We model the distribution of the factor errors, $\varepsilon_{i,t}$ with t - distribution with variance $\exp(h_{i,t})$ and ν degrees freedom. We note that t -distribution with ν degrees of freedom is essentially a scale mixture of normal distribution and hence

$$\varepsilon_{i,t} = \gamma_t^{-1/2} \exp\left(\frac{h_{i,t}}{2}\right) \epsilon_t \quad (\text{A.7})$$

where ϵ_t follows a standard normal distribution. When γ_t follows a Gamma distribution with $\Gamma(\frac{\nu}{2}, \frac{\nu}{2})$, then $\varepsilon_{i,t}$ follows a Student's t -distribution with ν degrees of freedom, see Geweke (1993), Geweke (2005) for textbook expositions and Curdia *et al.* (2014) for an application in the context of structural macroeconomic models.