A Dynamic Nelson-Siegel Yield Curve Model with Markov Switching

Jared Levant[ǂ]
Jun Ma[§]

November 7, 2013

Abstract

This paper proposes a model to better capture persistent regime changes in the interest rates of the US term structure. While the previous literature on this matter proposes that regime changes in the term structure are due to persistent changes in the conditional mean and volatility of interest rates we find that changes in a single parameter of the model we use better models regime changes. Furthermore, we investigate if the effects of macroeconomic phenomena such monetary policy, inflation expectations, and real business activity differ according to the particular regime realized for the term structure. Our results indicate that in periods of low interest rates, monetary policy and real business activity have a greater effect on the longer maturities of the yield curve than in high interest rate regimes. In those periods of high interest rate regimes, inflation expectations have a greater effect in yield determination for longer maturities. (JEL: C51, E43)

Keywords: Term Structure, Regime Shifts, Nelson–Siegel model, State-Space model, Kalman Filter, Kim Filter, Bootstrap

[ǂ]Corresponding author: Culverhouse College of Commerce & Business Administration, University of Alabama, Tuscaloosa, AL, USA, jlevant@crimson.ua.edu
[§] Culverhouse College of Commerce & Business Administration, University of Alabama, Tuscaloosa, AL, USA, jma@cba.ua.edu

1. INTRODUCTION

The yield curve often contains useful information about the real economic activity and inflation. For example, the level factor (the long-term yield-to-maturity) is often argued to be closely related with the inflation expectation, while the steepness or the slope factor (the long-term yield-to-maturity minus the short-term yield-to-maturity) has been shown to vary with the business cycles and also be heavily influenced by the monetary policy. The most recent monetary policies, such as the operation twist conducted by the Federal Reserve Bank in an attempt to lower the long-term interest rate and raise the short-term rate, directly work on the yields curve and serve as a great example of how the yield curve, instead of just one single policy rate–federal funds rate–is expected to have a significant impact on the economy. As such, it is important to correctly model the yield curve to understand better its interactions with the business cycles, and the monetary policy transmission mechanism through its impacts on the yield curve.

One popular approach to modeling the yields curve in literature is to impose no-arbitrage conditions and derive the yields curve based on latent factors. The majority of research in this area works on the affine class of models for the purpose of tractability. Duffie and Kan (1996) and Dai and Singleton (2000) work out a general class of the affine term structure model which encompasses some of the early models such as Vasicek (1977) and Cox, Ingersoll, and Ross (1985). Despite being appealing theoretically, these models in general forecast poorly, likely due to their restrictive nature, as pointed out by Duffee (2002).

A second class of models rectified the short comings of the first class in a novel way. Employing the relationship from expectations theory, Nelson and Siegel (1987, NS) was able to model forward rates directly with a three latent factor model and derive the yield curve. The three latent factors represent the level, slope, and curvature of the yield curve. Unlike the no-arbitrage affine models, the NS model greatly improved forecasting across bond maturities and has become very popular, in particular among the central banks.[1] Moreover, recent work by

[1] See, for example, the BIS (2005) report that points out the central banks in a number of countries including France, Germany, and Spain all use the NS model to model the yield curve.

Coroneo, Nyholm and Vidova-Koleva (2011) find that the NS model is close to being arbitrage-free when applied to the US market, although it does not explicitly impose these restrictions.[2]

Diebold and Li (2006) extended the NS model to the time dimension by allowing the three latent factors of the NS model to be time-varying. This dynamic Nelson-Siegel (DNS) model's forecast has been shown to outperform vector autoregressive models and dynamic error correction models. Diebold, Rudebusch, and Aruoba (2006) formulate the DNS model into a state-space framework, employing the Kalman filter and maximum likelihood estimation to estimate the three latent factors together with some observed macroeconomic factors. Diebold, Li, and Yue (2008) further model the yield curves in a set of countries and extract the global yield factors that appear to explain a large portion of the global yield curves movements. The re-interpretation of the NS model by Diebold and Li (2006) emphasizes the factor structure of the NS model. Their extension of the original NS model to the dynamic NS model (DNS) clearly shows that the NS model is essentially a particular factor model that captures the whole yield curve movement through a few latent factors with a set of specified loading parameters. In this way the DNS model is closely related with the Dynamic Factor Model (DFM) that has often been applied to macroeconomic and finance data, such as the work by Stock and Watson (1991) that extracts an economic coincident index from a set of economic variables using the DFM.

The interest rate dynamics of the term structure has been subject to a number of structural breaks historically. Although some of these breaks or regime changes are results of obvious changes in monetary policy as in the Volcker era and obvious changes in business cycle conditions such as the oil supply shock of the 1970s, there are also other regime changes that can be traced to occurring business cycle fluctuations such as troughs and peaks and often indirectly observed changes in the financial markets. As a result it is important to capture these regime changes in order to obtain a more accurate estimation of the term structure, which in turn can give not only a better understanding of the past and current economy but also a better prediction of the future economy. Recently a growing literature has also started to examine the consequence of structural breaks on the estimation of the DFM. For example, Breitung and Eickmeier (2011) suggest using the *LM* test to search for the exogenous type of breaks of the

[2] Svensson (1994) introduced a second curvature factor, and thus a second loading parameter λ, to the NS framework to better estimate longer termed maturities, a noted weakness of the NS model. Christensen, Diebold, and Rudebusch (2009) included a second slope factor as well as a second curvature factor to derive an arbitrage-free class of NS models.

loading parameters in the DFM. Stock and Watson (2008) investigates the forecasting performance of the DFM in the presence of the parameter instability.

In context of the NS model, Koopman, Mallee, and Van der Wel (2010) include the loading parameter λ as a time-varying latent factor to be estimated along with the three time-varying latent factors via the extended Kalman filter, as a way to allow the time variation in the parameters. They also introduce time-varying volatility in the form of a GARCH process into the DNS framework as a way of relaxing a specification assumption made by Bianchi, Mumtaz, and Surico (2006) that the factor loadings are also appropriate weights for the term structure's volatility. Wong, Lucia, Price and Startz (2011) study the connection of the yield curves in US and Canada, and identify an exogenous structural break in the NS model that reveals a weaker correlation between the yield curves in these two countries after Canada changed its monetary policy and switched to the explicit inflation targets in 1991. While Startz and Tsang (2010) incorporate Markov regime switching into an unobserved components model of the yield curve to account for regime changes of the yield curve. As an alternative modeling approach to the exogenous type of breaks, Markov regime switching proposed in Hamilton (1989) has the advantage that the underlying breaks can be reoccurring and stochastic in nature. Markov regime switching has been successfully introduced to the DFM by Chauvet (1998), and Kim and Nelson (1998) that generalize the work of Stock and Watson (1991) to allow Markov regime switching in extracting an economic coincident index from a set of macroeconomic variables.

We contribute to the literature by introducing and thoroughly evaluating regime-switching factor loadings and regime-switching volatility in the dynamic Nelson-Siegel model. In our models, regimes are characterized by a latent Markov switching component—the fourth latent factor in our models. This factor dictates which state drives the system's dynamics. When we apply a Markov switching component to the loading parameter, the slope and curvature factor loadings will assume distinct values in each state for each maturity according to the value of the loading parameter in each state. The Kalman filter (KF) can efficiently extract the level, slope, and curvature factors while the Kim (1994) algorithm allows us to extract the states. Next, we introduce regime-switching volatility. Following the assumption of Bianchi, Mumtaz and Surico (2006) that the factor loadings are appropriate weights for the term structure's volatility, we apply a Markov switching component to the factor disturbances. We again utilize the KF and the Kim filter to estimate our model's latent factors. Comparisons between the models are made by presenting goodness-of-fit statistics and AIC/BIC values. We finally implement likelihood ratio tests to investigate if our models are statistically different from the baseline dynamic Nelson-Siegel model. The root mean square error analysis shows the model with the loading parameter switching estimates yields the best across the short, medium, and long maturity ranges and in terms of overall fit. This model also gives the minimum AIC/BIC values of all models under consideration. Both models are found to be statistically different from the baseline model.

A number of papers investigate the role of the loading parameter and volatility on yield curve estimation. We have already discussed Koopman et al. (2010) and Bianchi, Mumtaz, and Surico (2006). Yu and Salyards (2009) and Yu and Zivot (2011) apply the DNS model to modeling corporate bond yields. The findings from both papers suggest that the optimal λ changes as one goes from modeling investment to speculative grade bonds. These results corroborate our general findings. The paper is organized as follows. Section 2 describes the dynamic Nelson-Siegel model, Kalman filter and the Kim algorithm. Section 3 describes the data. Section 4 presents the results of the various models and section 5 concludes.

2. MODELS and ESTIMATION

In this section we introduce our baseline model, the dynamic Nelson-Siegel (DNS) model. The appeal of this model lies in its extension to the time dimension. Also, the formulation lowers the coherence between the slope and curvature factor loadings. This diminished correlation between factor loadings aids any statistical analysis involving the NS framework. In addition to introducing the DNS model we introduce our regime-switching models and the estimation technique used.

2.1 The Dynamic Nelson-Siegel Model

The Diebold and Li (2006) factorization of the NS model is given by

$$\mathrm{y_t(m)} = \mathrm{y_t}(\mathrm{m}; \boldsymbol{F}_t, \lambda) = \mathrm{L_t} + \mathrm{S_t}\frac{1-\mathrm{e}^{-\lambda \mathrm{m}}}{\lambda \mathrm{m}} + \mathrm{C_t}\left(\frac{1-\mathrm{e}^{-\lambda \mathrm{m}}}{\lambda \mathrm{m}} - \mathrm{e}^{-\lambda \mathrm{m}}\right) \tag{1}$$

where $\boldsymbol{F}_t = (\mathrm{L_t}, \mathrm{S_t}, \mathrm{C_t})'$, for given time t, maturity m, and constant λ, the factor loading parameter. This is the baseline DNS model in our analysis.

The shape of the yield curve comes from the factor loadings and their respective weights in F_t. From Equation (1), the factor loading associated with $\mathrm{L_t}$ is assumed to be 1 independent of maturity and therefore influences short, medium, and long-term interest rates equally. The loading factors for $\mathrm{S_t}$ and $\mathrm{C_t}$ depend on both maturity and the loading parameter. For a given t, the slope factor loading converges to one as $\lambda \downarrow 0$ (or $m \downarrow 0$) and converges to zero as $\lambda \to \infty$ (or $m \to \infty$). The curvature factor loading converges to zero as $\lambda \downarrow 0$ (or $m \downarrow 0$) and as $\lambda \to \infty$ (or $m \to \infty$) for a given t.

Since we are interested in the loading parameter's effect on yields, we use the limit analysis above to understand the asymptotic behavior of the yield curve. The yield curve converges to $L + S$ as $\lambda \downarrow 0$ and converges to L as $\lambda \to \infty$ for a given t. These limiting values indicate that without the loading parameter the yield curve is flat and with extreme values for the loading parameter the yield curve would become flat. So "reasonable" values for λ are responsible for the wide range of non-flat yield curve shapes within an NS framework.

2.2 DNS Model Estimation

We adopt Diebold, Rudebusch, and Aruoba (2006) state-space framework to model each variant of the NS model in this paper. Our measurement equation models the time-series process of the yields according to the latent factors and takes the form

$$\begin{pmatrix} \mathrm{y_t(m_1)} \\ \mathrm{y_t(m_2)} \\ \vdots \\ \mathrm{y_t(m_N)} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda m_1}}{\lambda m_1} & \frac{1-e^{-\lambda m_1}}{\lambda m_1} - e^{-\lambda m_1} \\ 1 & \frac{1-e^{-\lambda m_2}}{\lambda m_2} & \frac{1-e^{-\lambda m_2}}{\lambda m_2} - e^{-\lambda m_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda m_N}}{\lambda m_N} & \frac{1-e^{-\lambda m_N}}{\lambda m_N} - e^{-\lambda m_N} \end{pmatrix} \begin{pmatrix} \mathrm{L_t} \\ \mathrm{S_t} \\ \mathrm{C_t} \end{pmatrix} + \begin{pmatrix} \varepsilon_\mathrm{t}(\mathrm{m_1}) \\ \varepsilon_\mathrm{t}(\mathrm{m_2}) \\ \vdots \\ \varepsilon_\mathrm{t}(\mathrm{m_N}) \end{pmatrix} \tag{2}$$

or expressed in matrix notation as

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}(\lambda)\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim MN(0, \boldsymbol{\Sigma}_\varepsilon), t = 1, \dots, T, \tag{3}$$

with $\boldsymbol{y}_t$ representing the $N \times 1$ vector of yields, $N \times 3$ factor loading matrix $\boldsymbol{\Lambda}(\lambda)$, 3×1 latent factor vector $\boldsymbol{F}_t$, and $N \times 1$ yield disturbance vector $\boldsymbol{\varepsilon}_t$ (or so-called measurement errors of the yields). The diagonal structure of $\boldsymbol{\Sigma}_\varepsilon$ implies that measurement errors across maturities of $\boldsymbol{y}_t$ are uncorrelated and is a fairly standard assumption in the literature. The transition equation, which models the time series process of the latent factors, can be expressed by the vector autoregressive (VAR) process

$$\begin{pmatrix} \mathrm{L_t} - \mu_{L_t} \\ \mathrm{S_t} - \mu_{S_t} \\ \mathrm{C_t} - \mu_{C_t} \end{pmatrix} = \begin{pmatrix} \mathrm{a_{11}} & 0 & 0 \\ 0 & \mathrm{a_{22}} & 0 \\ 0 & 0 & \mathrm{a_{33}} \end{pmatrix} \begin{pmatrix} \mathrm{L_{t-1}} - \mu_{L_t} \\ \mathrm{S_{t-1}} - \mu_{S_t} \\ \mathrm{C_{t-1}} - \mu_{C_t} \end{pmatrix} + \begin{pmatrix} \eta_{L_t} \\ \eta_{S_t} \\ \eta_{C_t} \end{pmatrix}$$

which can be equivalently expressed in matrix notation as

$$\boldsymbol{F}_t = (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{F}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim MN\left(0, \boldsymbol{\Sigma}_\eta\right), t = 1, \dots, T, \tag{4}$$

with 3×1 mean vector $\boldsymbol{\mu}$, 3×3 coefficient matrix $\boldsymbol{A}$, and 3×1 factor disturbance matrix $\boldsymbol{\Sigma}_\eta$. Christensen et al. (2011) show that the off-diagonal elements of the $\boldsymbol{A}$ matrix are not statistically relevant to modeling the term structure. The assumption that the deviations to the NS factors are uncorrelated is not standard in the literature. But Diebold et al. (2006) conclude that the off-diagonal elements are marginally significant and the point estimates and standard errors of the $\boldsymbol{A}$ matrix are little changed when estimating $\boldsymbol{\Sigma}_\eta$.

Since the DNS state space model is linear in latent factors, we are able to use the Kalman filter to estimate the latent factors conditional on past and contemporaneous observations of the yields. The Kalman filter procedure is carried out recursively for $t = 1, \dots, T$ with initial values for the latent factors and their variances being the unconditional mean and unconditional variance, respectively. If we define $\boldsymbol{f}_{t|t}$ as the minimum mean square linear estimator (MMSLE) of $\boldsymbol{F}_t$ and $\boldsymbol{v}_{t|t}$ as the mean square error (MSE) matrix, then $\boldsymbol{f}_{1|0} = \boldsymbol{\mu}$ and $\boldsymbol{v}_{1|0} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{\Sigma}_\eta$. With observation $\boldsymbol{y}_t$ and initial values$\boldsymbol{f}_{1|0}$ and $\boldsymbol{v}_{1|0}$ available, the KF updates the values for $\boldsymbol{f}_{t|t}$ and $\boldsymbol{v}_{t|t}$ using the equations

$$\boldsymbol{f}_{t|t} = \boldsymbol{f}_{t|t-1} + \boldsymbol{K}_t \boldsymbol{e}_{t|t-1}, \tag{5}$$

$$\boldsymbol{v}_{t|t} = \boldsymbol{v}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{\Lambda}(\lambda) \boldsymbol{v}_{t|t-1}, \tag{6}$$

where $\boldsymbol{e}_{t|t-1} = \boldsymbol{y}_t - \boldsymbol{\Lambda}(\lambda)\boldsymbol{f}_{t|t-1}$ is the predicted error vector, $\boldsymbol{ev}_{t|t-1} = \boldsymbol{\Lambda}(\lambda)\boldsymbol{v}_{t|t-1}\boldsymbol{\Lambda}(\lambda)' + \boldsymbol{\Sigma}_\varepsilon$ is the predicted error variance matrix and $\boldsymbol{K}_t = \boldsymbol{v}_{t|t-1}\boldsymbol{\Lambda}(\lambda)'\boldsymbol{ev}_{t|t-1}^{-1}$ is the Kalman gain matrix. The next period $t+1$ MMSLE of the latent factors and associated variance matrix conditional on yields $\boldsymbol{y}_1, \dots, \boldsymbol{y}_t$ are governed by the prediction equations

$$\boldsymbol{f}_{t|t-1} = (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{f}_{t-1|t-1} \tag{7}$$

$$\boldsymbol{v}_{t|t-1} = \boldsymbol{A}\boldsymbol{v}_{t-1|t-1}\boldsymbol{A}' + \boldsymbol{\Sigma}_\eta. \tag{8}$$

Denote $\boldsymbol{\theta}$ as the system parameter vector. The parameters to be estimated via numerical maximum likelihood estimation are $\boldsymbol{\theta}_{DNS} = \left\{\boldsymbol{A}_{ij}, \boldsymbol{\Sigma}_{\varepsilon_{ij}}, \boldsymbol{\Sigma}_{\eta_{ij}}, \boldsymbol{\mu}, \lambda\right\}$. We represent the likelihood function as

$$\ell(\boldsymbol{\theta}) = -\frac{NT}{2} log 2\pi - \frac{1}{2}\sum_{t=1}^{T} \log |\boldsymbol{ev}_t| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{e}_t'(\boldsymbol{ev}_t)^{-1}\boldsymbol{e}_t. \tag{9}$$

The function $\ell(\boldsymbol{\theta})$ is evaluated by the Kalman filter through a quasi-Newton optimization method for the purposes of maximization without inverting the Hessian matrix of this 28-parameter system.

The values for $\boldsymbol{f}_{t|t}$ and $\boldsymbol{v}_{t|t}$ from the last iteration of the KF are used as initial values in the recursive algorithm to obtain smoothed values of the unobserved factors. Iterating the following two equations backwards for $t = T-1, T-2, \ldots 1$, gives the smoothed estimates:

$$\boldsymbol{f}_{t|T} = \boldsymbol{f}_{t|t} + \boldsymbol{v}_{t|t}\boldsymbol{\Lambda}(\lambda)'\boldsymbol{v}_{t+1|t}^{-1}(\boldsymbol{f}_{t+1|T} - \boldsymbol{\Lambda}(\lambda)\boldsymbol{f}_{t|t} - \boldsymbol{\mu}), \tag{10}$$

$$\boldsymbol{v}_{t|T} = \boldsymbol{v}_{t|t} + \boldsymbol{v}_{t|t}\boldsymbol{\Lambda}(\lambda)'\boldsymbol{v}_{t+1|t}^{-1}\big(\boldsymbol{v}_{t+1|T} - \boldsymbol{f}_{t+1|t}\big)\boldsymbol{v}_{t+1|t}^{-1}{}'\boldsymbol{\Lambda}(\lambda)\boldsymbol{v}_{t|t}. \tag{11}$$

These smoothed estimates provide a more accurate inference on $\boldsymbol{f}_t$ because it uses more information from the system than the filtered estimates.

2.3 The DNS Model with Regime-Switching Loading Parameter

In sub-section 2.1 we established that λ determines the shape of the yield curve. Thus changes in interest rate levels are determined by λ, given other factors. Realizing that keeping λ fixed across the sample period may be a source of model mis-specification in the literature (see Diebold and Li (2006) Diebold et. al (2006), and Xiang and Zhu (2013)), Koopman et al. (2010) treat λ as a time-varying latent factor of the model to be estimated in the same fashion as the latent NS factors of the model.

We model λ as a regime-switching parameter that influences interest rate levels according to the realized state. We assume the term structure follows a two-state regime switching process for computational tractability of our model. Investigating ex-post real interest rates, Garcia and Perron (1996) assume interest rates follow a three-state regime switching process. And using a reversible jump Markov chain Monte Carlo (RJMCMC) procedure, Xiang and Zhu (2013) estimate two distinct regimes for the term structure.

We propose to treat the loading parameter λ as a regime switching parameter solely determined by the realized state of the yields, S_t. The latent Markov component S_t is governed by a two-state Markov process and we denote the states simply as 0 or 1 corresponding to the term structure being in the low or high regime respectively. The loading matrix $\boldsymbol{\Lambda}(\lambda)$ in Equation (3) is replaced by $\boldsymbol{\Lambda}(\lambda_{S_t})$ and the resulting measurement equation is $\boldsymbol{y}_t = \boldsymbol{\Lambda}(\lambda_{S_t})\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$. Take note that since we are not including λ in $\boldsymbol{F}_t$, the observation vector of yields is still linear with respect to our latent factor vector. This new measurement equation along with our transition equation from Equation (4) constitutes the DNS-MSL model for estimation with regime-switching the loading parameter.

2.4 The DNS Model with Regime-Switching Volatility for Factors

In most of the empirical literature on term structure modeling, a constant volatility is assumed in the time-series of interest rates. Like modeling the DNS model with constant loading parameter, a constant volatility over time may be a source of model misspecification for estimating the term structure. A few papers investigate time-varying volatility in the context of the DNS model. Bianchi et al. (2009) employ a VAR augmented with NS factors and macro-factors featuring time-varying coefficients and stochastic volatility. Koopman et al. (2010) estimate yield disturbances according to a GARCH specification to introduce a time-varying variance.

We modify the DNS model by introducing regime-switching factor disturbances by applying a hidden Markov switching component to the factor disturbances in the transition equation. Equation (3) represents the measurement equation of this model and after replacing $\boldsymbol{\eta}_t$ with $\boldsymbol{\eta}_{S_t}$ the new transition equation is

$$\boldsymbol{F}_t = (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{F}_{t-1} + \boldsymbol{\eta}_{S_t}, \qquad \boldsymbol{\eta}_{S_t} \sim MN\left(0, \boldsymbol{\Sigma}_{\eta,S_t}\right), t = 1, \dots, T, S_t = 0,1.$$

The state-space model comprising the measurement equation from equation (3) and volatility switching transition equation with $\boldsymbol{\eta}_{S_t}$ substituted into equation (4) will be referred to as the DNS-MSV model.

2.5 Estimation Based on the Kim Filter

In this sub-section we will show that the Kim filter allows for efficient estimation of parameters through the KF and accurate inference of the realized states through a methodology developed by Hamilton (1989, 1990). But before we outline filter, the DNS-MSL state-space model can be represented in its entirety as

$$\begin{aligned}
&\boldsymbol{y}_t = \boldsymbol{\Lambda}(\lambda_{S_t})\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, && \boldsymbol{\varepsilon}_t \sim MN(0, \boldsymbol{\Sigma}_\varepsilon), t = 1, \dots, T,\\
&\boldsymbol{F}_t = (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{F}_{t-1} + \boldsymbol{\eta}_t, && \boldsymbol{\eta}_t \sim MN\left(0, \boldsymbol{\Sigma}_\eta\right), t = 1, \dots, T, \qquad (12)\\
&\lambda_{S_t} = \lambda_0(1-S_t) + \lambda_1 S_t, && S_t = 0,1,
\end{aligned}$$

and the DNS-MSV model in its entirety is

$$\begin{aligned}
&\boldsymbol{y}_t = \boldsymbol{\Lambda}(\lambda)\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, && \boldsymbol{\varepsilon}_t \sim MN(0, \boldsymbol{\Sigma}_\varepsilon), t = 1, \dots, T,\\
&\boldsymbol{F}_t = (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{F}_{t-1} + \boldsymbol{\eta}_{S_t}, && \boldsymbol{\eta}_t \sim MN\left(0, \boldsymbol{\Sigma}_{\eta S_t}\right), t = 1, \dots, T, \qquad (13)
\end{aligned}$$

$$\boldsymbol{\eta}_{S_t} = \eta_0(1 - S_t) + \eta_1 S_t, \qquad S_t = 0,1,$$

where for both models the transition probabilities between states are governed by the entries of the matrix

$$\begin{pmatrix} p_{00} = p & p_{01} = 1 - p \\ p_{10} = 1 - q & p_{11} = q \end{pmatrix}$$

where $p_{ij} = \Pr[S_t = j | S_{t-1} = i]$ with $\sum_{j=0}^{1} p_{ij} = 1$ for all i.

The estimation of the parameters of the model according to the Kim filter is very similar to the KF procedure explained for the non-switching case. Recall the latent factors for the DNS-MSL and DNS-MSV models are the NS factors and the unobserved state, S_t. We initialize the NS factors and their variances as in the non-switching case. To initialize the unobserved state, S_t, we need $\Pr[S_0 = j | \psi_0]$ where $j = 0,1$ and ψ_t refers to information up to time t. This expression is the steady state or unconditional probability of being in the low regime which is given by the formulas

$$\pi_0 = \Pr[S_0 = 0 | \psi_0] = \frac{1-p}{2-p-q}, \tag{14}$$

$$\pi_1 = \Pr[S_0 = 1 | \psi_0] = \frac{1-q}{2-p-q}, \tag{15}$$

where p and q are defined in the above transition probability matrix. Given realizations of the NS factors at t and $t-1$ when $S_{t-1} = i$ and $S_t = j$, the KF can be expressed as

$$\boldsymbol{f}_{t|t}^{(i,j)} = \boldsymbol{f}_{t|t-1}^{(i,j)} + \boldsymbol{K}_t^{(i,j)} \boldsymbol{e}_{t|t-1}^{(i,j)}, \tag{16}$$

$$\boldsymbol{v}_{t|t}^{(i,j)} = \boldsymbol{v}_{t|t-1}^{(i,j)} - \boldsymbol{K}_t^{(i,j)} \boldsymbol{\Lambda}(\lambda) \boldsymbol{v}_{t|t-1}^{(i,j)}, \tag{17}$$

$$\boldsymbol{e}_{t|t-1}^{(i,j)} = \boldsymbol{y}_t - \boldsymbol{\Lambda}(\lambda) \boldsymbol{f}_{t|t-1}^{(i,j)} \tag{18}$$

$$\boldsymbol{ev}_{t|t-1}^{(i,j)} = \boldsymbol{\Lambda}(\lambda) \boldsymbol{ev}_{t|t-1}^{(i,j)} \boldsymbol{\Lambda}(\lambda)' + \boldsymbol{\Sigma}_\varepsilon \tag{19}$$

$$\boldsymbol{f}_{t|t-1}^{(i,j)} = (\boldsymbol{I} - \boldsymbol{A}) \boldsymbol{\mu}_j + \boldsymbol{A} \boldsymbol{f}_{t-1|t-1}^{(i,j)} \tag{20}$$

$$\boldsymbol{v}_{t|t-1}^{(i,j)} = \boldsymbol{A} \boldsymbol{v}_{t-1|t-1}^{(i,j)} \boldsymbol{A}' + \boldsymbol{\Sigma}_\eta. \tag{21}$$

where $\boldsymbol{K}_t^{(i,j)} = \boldsymbol{v}_{t|t-1}^{(i,j)} \boldsymbol{\Lambda}(\lambda)' \left(\boldsymbol{ev}_{t|t-1}^{(i,j)} \right)^{-1}$ is the Kalman gain

The efficiency of the Kim filter arises from collapsing the (2×2) posteriors $\boldsymbol{f}_{t|t}^{(i,j)}$ and $\boldsymbol{v}_{t|t}^{(i,j)}$ into two single-state posteriors

$$\boldsymbol{f}_{t|t}^{j} = \frac{\sum_{i=0}^{1} \Pr[S_{t-1} = i, S_t = j | \psi_t] \boldsymbol{f}_{t|t}^{(i,j)}}{\Pr[S_t = j | \psi_t]}, \tag{22}$$

and

$$\boldsymbol{v}_{t|t}^{j} = \frac{\sum_{i=0}^{1} \Pr[S_{t-1}=i, S_t=j|\psi_t]\{\boldsymbol{v}_{t|t}^{(i,j)} + (\boldsymbol{f}_{t|t}^{(j)} - \boldsymbol{f}_{t|t}^{(i,j)})(\boldsymbol{f}_{t|t}^{(j)} - \boldsymbol{f}_{t|t}^{(i,j)})'\}}{\Pr[S_t=j|\psi_t]}, \tag{23}$$

by taking weighted averages over states at $t-1$. Following Hamilton (1989, 1990), the Kim (1994) filter is a consequence of Bayes' theorem which we can use to get the previous single-state posteriors results. Starting with the joint distribution of our states, we have

$$\begin{aligned}\Pr[S_t = j, S_{t-1} = i|\psi_t] &= \frac{\Pr[\boldsymbol{y}_t, S_t=j, S_{t-1}=i|\psi_{t-1}]}{\Pr[\boldsymbol{y}_t|\psi_{t-1}]} \\ &= \frac{f(\boldsymbol{y}_t|s_t=j, s_{t-1}=i, \psi_{t-1}) \times \Pr[S_t=j, S_{t-1}=i|\psi_{t-1}]}{\Pr[\boldsymbol{y}_t|\psi_{t-1}]} \end{aligned} \tag{24}$$

The two terms in the numerator and the probability in the denominator can be put in terms of known quantities from our estimation model. The conditional density $f(\boldsymbol{y}_t|s_{t-1} = i,\ s_t = j, \psi_{t-1})$ is obtained based on the prediction error decomposition:

$$f(\boldsymbol{y}_t|S_{t-1} = i,\ S_t = j, \psi_{t-1}) = (2\pi)^{-\frac{N}{2}}|\boldsymbol{ev}_{t|t-1}^{(i,j)}|exp\left\{-\frac{1}{2}\boldsymbol{e}_{t|t-1}^{(i,j)}{}'\left(\boldsymbol{ev}_{t|t-1}^{(i,j)}\right)^{-1}\boldsymbol{e}_{t|t-1}^{(i,j)}\right\}$$

and

$$\Pr[S_t = j, S_{t-1} = i|\psi_{t-1}] = \Pr[S_t = j|S_{t-1} = i] \times \Pr[S_{t-1} = i|\psi_{t-1}]$$

where $\Pr[S_t = j|S_{t-1} = i]$ is the transition probability. The terms in the numerator are now in known terms. The denominator, $\Pr[\boldsymbol{y}_t|\psi_{t-1}]$, can be expressed as

$$\Pr[\boldsymbol{y}_t|\psi_{t-1}] = \sum_{j=0}^{1}\sum_{i=0}^{1} \Pr[\boldsymbol{y}_t, S_t = j, S_{t-1} = i|\psi_{t-1}].$$

Finally, summing over state i we get our single state posterior

$$\Pr[S_t = j|\psi_t] = \sum_{i=0}^{1} \Pr[S_t = j, S_{t-1} = i|\psi_t]. \tag{25}$$

From the filter we obtain the density of $\boldsymbol{y}_t$ conditional on past information ψ_{t-1}, $t = 1,2,\dots,T$. We can now calculate maximum likelihood estimates from the approximate log likelihood function

$$\ell(\boldsymbol{\theta}) = ln[f(y_1, y_2, \dots, y_T)] = \sum_{t=1}^{T} \ln(\Pr[\boldsymbol{y}_t|\psi_{t-1}]). \tag{26}$$

Because these are switching models, the parameter vector set for both are going to be have more parameters estimated than the non-switching model: $\boldsymbol{\theta}_{DNS-MSL} = \{\boldsymbol{A}_{ij}, \boldsymbol{\Sigma}_{\varepsilon_{ij}}, \boldsymbol{\Sigma}_{\eta_{ij}}, \boldsymbol{\mu}, \lambda_0, \lambda_1\}$ and $\boldsymbol{\theta}_{DNS-MSV} = \{\boldsymbol{A}_{ij}, \boldsymbol{\Sigma}_{\varepsilon_{ij}}, \boldsymbol{\Sigma}_{\eta_{S_{t_{ij}}}}, \boldsymbol{\mu}, \lambda_0\}$.

Once we have finished calculating the maximum of $\ell(\boldsymbol{\theta})$, all parameters have been estimated and we can get inferences on S_t and $\boldsymbol{f}_t$ conditional on all the information in the sample: $Pr[S_t = j|\psi_t]$ and $\boldsymbol{f}_{t|T}$ for $t = 1,2,\dots,T$. Instead of incrementing to the end of the sample as in the KF, to obtain smoothed probabilities and factors we increment from the end of the sample to the beginning, gathering all information along the way. So for $t = T-1, T-2 \ldots, 1$ we can approximate the smoothed joint probability

$$\Pr[S_t = j, S_{t+1} = k|\psi_T] \approx \Pr[S_{t+1} = k|\psi_T] \times \Pr[S_t = j|\psi_t]$$
$$= \frac{\Pr[S_{t+1}=k|\psi_T] \times \Pr[S_t=j|\psi_t] \times \Pr[S_t=j, S_{t+1}=k|S_t=j]}{\Pr[S_{t+1}=k|\psi_t]} \qquad (27)$$

and probability

$$\Pr[S_t = j|\psi_T] = \sum_{k=0}^{1} \Pr[S_t = j, S_{t+1} = k|\psi_T] \qquad (28)$$

These probabilities are used as weights in weighted averages to collapse the $(M \times M)$ elements of $\boldsymbol{f}_{t|T}^{(j,k)}$ and $\boldsymbol{v}_{t|T}^{(j,k)}$ into M where $M = 2$ for our model. These weighted averages over S_{t+1} are

$$\boldsymbol{f}_{t|T}^{j} = \frac{\sum_{k=0}^{1} \Pr[S_t=j, S_{t+1}=k|\psi_T] \boldsymbol{f}_{t|T}^{(j,k)}}{\Pr[S_t=j|\psi_T]} \qquad (29)$$

and

$$\boldsymbol{v}_{t|T}^{j} = \frac{\sum_{k=0}^{1} \Pr[S_t=j, S_{t+1}=k|\psi_T] \{\boldsymbol{v}_{t|T}^{(j,k)} + (\boldsymbol{f}_{t|T}^{j} - \boldsymbol{f}_{t|T}^{(j,k)})(\boldsymbol{f}_{t|T}^{j} - \boldsymbol{f}_{t|T}^{(j,k)})'\}}{\Pr[S_t=k|\psi_T]} \qquad (30)$$

Taking a weighted average over the states at time t we get an expression for the smoothed factors

$$\boldsymbol{f}_{t|T} = \sum_{j=0}^{1} \Pr[S_t = j|\psi_T] \, \boldsymbol{f}_{t|T}^{j}. \qquad (31)$$

This completes the Kim filter. Further details and justifications can be found in Kim and Nelson (1999).

3. DATA

We use end-of-month, bid-ask averages for U.S. Treasury yields from January 1970 through December 2000. Diebold and Li (2006) convert the unsmoothed Fama-Bliss (1987) forward rates given by CRSP to unsmoothed Fama-Bliss zero rates for the following eighteen maturities: 1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 month. The data is kindly supplied by Francis Diebold. Figure 1 gives a 3-dimensional mesh plot of the term structure for the sample period and the maturities.

[Insert Figure 1 here]

There are two primary advantages for us to limit our study to this particular sample period: first, this Diebold's dataset is produced through consistent and careful cleaning procedures that remove microstructure noises and therefore provides us with a nice dataset to evaluate the proposed models; second, the most recent periods have featured an environment of extremely low nominal interest rates with the so-called zero lower bound constraint, which implies that the Gaussian type of term structure models typically encountered in the literature including the one implemented here would provide a poor approximation and thus calls for separate treatment that we leave for future studies.

Table 1 reports the means, standard deviations, and autocorrelations across maturities for the yields with maturities of 1, 12, and 30 months. The summary statistics show the average yield curve is upward sloping –a reflection of the risk premium inherent in longer maturities. The volatility is generally decreasing by maturity with the exceptions of the one-month being less volatile than the 3, 6, and 9-month bills and the 8-yr being less volatile than the 9-yr bond.

[Insert Table 1 here]

We also report the statistics for the empirical counterparts for the level, slope, and curvature factors. It is worth declaring which convention we adopt in calculating the empirical factors.[3] The empirical level factor is calculated as an average of the 1, 24, and 120 month maturities. The empirical slope factor is the difference between the 120 and 1 month maturities. Lastly, the empirical curvature factor is twice the 24 month maturity minus the sum of the 1 and 120 month.

4. EMPRICAL RESULTS

In this section we present the comparative test results for the DNS, DNS-MSL, and DNS-MSV models. Table 2 gives the correlations between the empirical factors and our estimated smoothed factors via the Kim filter for each model and Figures 3-5 show graphically the evolution of each factor and their empirical counterpart for each model. We identify a drop in correlation between the empirical curvature factor and smoothed curvature factor for the DNS-MSL model. The formulation of each empirical counterpart assumes no switching so it is not surprising that the factor which has two terms subject to switching would produce the lowest correlation.

Table 4 reports our in-sample root mean squared error (RMSE) values for the various models. Recall that the Kalman filter estimates measurement error parameters for each maturity.

---

[3] Some authors define empirical level as simply the observed long term maturity, which in our case would be y(120).

These parameter estimates are recorded as the diagonal terms of the covariance matrix of the measurement equation error. Taking these diagonal elements we are able to calculate the RMSE according to the formula

$$(\sum_1^T (y_t - \hat{y}_t)^2/T)^{1/2}$$

where T = 371 for all models. We find overall the DNS model yields the largest average RMSE and the DNS-MSL yields the smallest average RMSE. In terms of percentage changes, the DNS-MSL and the DNS-MSV decrease the in-sample average RMSE by 2.86 percent and 0.37 percent, respectively.

In addition to calculating the total average RMSE for all maturities, we calculate average RMSEs for the ranges of short, medium, and long-termed maturities. We limit the three maturity ranges to 6 maturities, i.e. short maturity range contains the 1, 3, 6, 9, 12 and 15m maturities, etc. The DNS-MSL decreases the average RMSEs for the short, medium, and long maturity range groups by 4.89 percent 2.47 percent, and 0.75 percent, respectively, the largest decreases for the respective groups. The improved modeling of the DNS-MSV model came with a decreased average RMSE for the short maturity range group by 1.30 percent but for the medium and long groups the average RMSE actually increased by 0.30 percent and 0.17 percent respectively over the DNS model. The DNS-MSV model is capturing the volatility of the short maturity bills but the bonds have less volatility so the model is over-identified for longer maturities resulting in greater loss of estimation efficiencies. These results support the case for applying a switching component to the DNS model. Furthermore, these results suggest switching λ leads to a noticeable improvement for the in-sample fit over the DNS model by better estimating longer termed maturities, a known deficiency of the DNS model.

The maximum log-likelihood value and AIC/BIC measures are calculated for each model and are given in Table 2. Our regime switching models show significant model improvement over the baseline DNS model. The DNS-MSL model achieves the greatest log-likelihood value and the smallest AIC and BIC values, further strengthening the conclusion that this model performs the best in-sample fit of all the models.

[Insert Table 2 here]

4.1 DNS Model

The baseline DNS model is estimated with parameter estimate values close to those of other DNS parameter estimates in the literature. The parameter estimates are listed in Table 2.

The estimate for the loading parameter λ is 0.080 with a standard error of 0.0035 while the estimated λ for Diebold, Rudebusch and Aruoba (2006) is 0.077. Using the Diebold and Li's interpretation of λ we are able to ascertain the maturity in which the loading on the curvature factor attains a maximum, henceforth referred to as implied maturities. Recall the loading on the curvature factor (CL) has functional form

$$CL = \frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}.$$

Taking the first order condition of CL with respect to the maturity, m, yields

$$CL_m = \frac{e^{-\lambda m}}{\lambda} + me^{-\lambda m} - \frac{1-e^{-\lambda m}}{m\lambda^2}.$$

Setting this nonlinear equation to zero and solving for m gives the maturity that the curvature loading reaches a maximum. The implied maturity of our CL is 22.4 months while the implied maturity in the DRA paper is 23.3 months.

Our estimated smooth factors have relatively high correlations to the empirical factors as shown in Table 3 and with the smoothed-empirical factor plots in Figure 3.

[Insert Table 3 here]

[Insert Figure 2 here]

Table 4 gives the RMSE calculations. The one and three-month rates are the most difficult to estimate and the medium-term maturities are fitted the best.

[Insert Table 4 here]

4.2 DNS-MSL Model

We introduce the first of our two regime switching models--the DNS-MSL model. Comparing the correlations of our estimated smoothed factors for this model and their respective empirical factors we find a drop in the correlations across the factors as compared with the baseline model. Specifically, the curvature factor experiences the largest decrease. This is further supported with a visual inspection of the third plot of the smoothed-empirical factor plots of Figure 3.

[Insert Figure 3 here]

There are two causes for this drastic decrease in the correlation of the smoothed curvature factor and its empirical counterpart in both switching models. First, the empirical factors are calculated under the assumption of no switching in yields. Therefore the factors responsible for capturing the switching we propose exists in the term structure –slope and curvature—should experience the largest decreases in correlation with empirical slope and curvature. Second, and more specifically for the curvature factor, the literature has shown that the curvature factor is highly volatile and thus may suffer from weak identification and therefore its estimation is the most tenuous of all the estimable factors. It is indeed the case that for each model we estimate the curvature factor has the highest volatility.

Our estimation results from the Kalman filter indicate the loading parameter λ is subject to a hidden Markov switching component. We estimate λ to be 0.055 and 0.153 for the low and high interest rate regimes, respectively. The implied maturities are 32.6 months and 11.7 months, respectively. Figure 5 shows the effect these values have on the slope and curvature factor loadings across maturities.

[Insert Figure 5 here]

In the first two plots we see a much faster decay of the slope and curvature loading factors in the high interest rate regime than in the low interest rate regime, therefore the slope and curvature factor loadings influence medium and long-term maturities less than in the low regime. This suggests that during periods when yields are relatively high, business cycle activity and monetary policy contribute less to yield determination in medium and long-term maturities than when yields are relatively low (see Figure 6 and the discussions of regimes below the figure). To put it another way, yields for medium and long-term maturities are determined more by long-term inflation expectations when yields are relatively high. Yu and Zivot (2011) find that differing patterns in the curvature factor of speculative-grade bonds suggest highly risky bonds are not sensitive to monetary policy.

From the third plot in Figure 5, we see that the slope loading factor is uniformly greater across maturities in the low regime than in the high regime. This shows definitively that the slope factor contributes more to yield determination over all maturities when yields are relatively low and thus monetary policy and business cycle activity play a more pivotal role in the level of yields. This was evident during Fed Chairman Greenspan's tenure in the Federal Reserve. In the last plot, the curvature loading factor is greater in the high regime than in the low regime for the one through 19-month maturities and therefore influences the yields of those maturities more so than in the low regime. For longer maturities in the high regime, the curvature loading factor decays quickly and is less of a factor in yield determination than in the low regime.

The DNS-MSL smoothed probability plot of Figure 6 shows the time-series of the unobserved state, S_t.

[Insert Figure 6 here]

The timing and duration of the regime changes in Figure 6 coincide very nicely with periods of high interest rates documented in the literature. In the early to mid-90s yields achieved their lowest levels in our sample period. This period corresponds to the Great Moderation where the level and volatility in interest rates decreased substantially. From the mid-90s to the late-90s we observe persistent regime changes. While interest rate levels were not as high as in the 1970s or 1980s, the yield levels of the mid-to-late 90s were high relative to the levels in the early-to-mid 90s. The loading parameter is capturing these relative changes in interest rate levels.

4.3 DNS-MSV Model

The second of our regime-switching models is the DNS-MSV model. As mentioned previously, authors investigating regime-switching in interest rates have consistently applied a hidden switching component to the conditional mean and volatility, simultaneously. After running pre-tests for Markov switching in the variaous model parameters, we find no significant switching in the conditional mean but we do find significant switching in the volatility parameters of each of the latent factors. From Table 2 we find the factor volatilities for the DNS-MSV model increase from 0.26, 0.33, and 0.66 in the low volatility regime to 0.50, 1.21, and 1.87, respectively, in the high volatility regime. These are relative increases of 92 percent, 267 percent, and 183 percent, respectively.

We estimate λ to be 0.081. Our λ estimate gives an implied maturity of 22.1 months. slightly smaller than the baseline's implied maturity of 22.4 months.

The timing of the regime changes in the volatility corresponds quite closely to terms structure volatility changes other papers have documented over the period from 1970 through 2000. Figure 7 plots the smoothed probabilities according to S_t.

[Insert Figure 7 here]

Figure 7 shows many abrupt but short-lived regime changes in the 1970s. But in the 1980s we see more persistent regime changes up until the latter half where we once again see distinct spikes. The term structure of the 1990s can be described as being in the low volatility regime throughout the decade. This dramatic change corresponds to the Great Moderation.

In accordance with the literature, we estimate a general model where both the decay parameter and factor disturbances are subject to switching, simultaneously. The model comparison tests show this model to be superior to the DNS-MSV results but inferior to the DNS-MSL results.

4.4 Are the DNS-MSL and DNS-MSV Models Statistically Different From the DNS Model?

We have established the DNS-MSL and DNS-MSV models outperform the DNS model with in-sample estimation. We now show the two models are statistically different from the DNS model. Since both the DNS-MSL and DNS-MSV models nest the DNS model, the likelihood ratio (LR) test is a good statistical testing candidate to address our issue. But we are unable to make accurate statistical inference using the asymptotic LR distribution for two reasons. First, the finite sample size may render the asymptotic theory less accurate in practice. Second, and more importantly, we encounter a nuisance parameter problem which renders asymptotic distributions used for testing nonstandard. Because of reason two, the classical optimality results of not only the LR test are invalid but also the Lagrange Multiplier (LM), Wald, and LR tests are rendered invalid. This issue has been addressed by Davies ((1977), (1987)), Andrews and Ploberger (1994), and Hansen ((1992), (1996)) among others. Although Andrews and Ploberger show the LR test is not an optimal test when deriving their new test in the presence of nuisance parameters, we follow Hansen's simulation methodology which does utilize the LR test.[4] The transition probabilities are the nuisance parameters in our model.

We outline the steps used to bootstrap the LR distribution for the comparison of the DNS and DNS-MSL models:

STEP 1: Obtain the max likelihood value (LLV_0) of the DNS model (null) and the max likelihood value (LLV_A) of the DNS-MSL model (alternative), using the real dataset. Calculate the $\widehat{LR}$ statistic.

STEP 2: Generate yields ($\hat{y}_0$) using the DNS model. Fit the DNS-MSL and DNS models to the yields ($\hat{y}_0$). Obtain the max likelihood value (LLV_A^*) for the DNS-MSL model and (LLV_0^*) for the DNS model. Calculate the new $\widehat{LR}_j^*$ using LLV_0^* and LLV_A^*.

[4] Our regime switching methodology is more closely related with Hansen (1992) when testing in the presence of nuisance parameters than Andrews and Ploberger (1992) which apply their test in the context of a single break in dynamics.

STEP 3: Using $\widehat{LR}$ and $\widehat{LR}_j^*$ compute a bootstrap critical value ($\hat{C}_\alpha^*$). For a test at level α, first sort the $\widehat{LR}_j^*$ from smallest to largest. Then calculate

$$\hat{C}_\alpha^* \simeq \widehat{LR}_{\alpha(B+1)}^*$$

where α represents your confidence level and B is the number of bootstraps. Repeat steps 2 and 3 B times and obtain $\{\widehat{LR}_j^*\}_1^B$.

STEP 4: Reject the null hypothesis if $\widehat{LR} > \hat{C}_\alpha^*$.

The steps are the same for deriving the LR distribution for the DNS and DNS-MSV comparison. The LR test statistic under the null of no switching is 467 for the DNS-MSL model and 410 for the DNS-MSV model. We perform 1000 bootstraps to derive a LR distribution. Figure 5 is a plot of the probability density for both models using a normal kernel function to smooth. Table 5 lists the critical vales for each model at the 10%, 5% and 1% confidence levels.

[Insert Table 5 here]

It is evident that the test statistic greatly exceeds all bootstrapped critical values so we are able to reject the null that the two models are statistically the same in both cases. This greatly enhances our stance that term structure modeling should take into account regime switching and that a model without regime switching is subject to omitted variable bias.

5. CONCLUSION

In this paper we investigate and model the parameter instability in the term structure using regime-switching dynamic Nelson-Siegel models. After applying a hidden Markov switching component to all of the model's parameters one at a time, we find that the factor loading parameter and latent factors' conditional volatilities show significant switching when allowed—not the conditional mean as noted in the literature. The model accounting for regime changes in volatility captured the timing of the volatility regimes associated with the oil price shock of the 1970s, monetary policy changes of the early 1980s and the period known as the Great Moderation. The model accounting for level changes in interest rates, i.e. switching the loading parameter modeled the regimes' correspondences to the oil price shock, the dis-inflationary policies under Fed chairman Paul Volker, and the aggressive monetary tightening and loosening under Fed chairman Alan Greenspan. The timing of the regime switches and their persistence according to smoothed probability plots strongly support our claim that the loading parameter is related to short-term monetary policy changes.

The model allowing a switching loading parameter yields smaller AIC/BIC values and produces smaller root mean squared error values for most of the individual maturities. The model also produced smaller RMSEs across maturity groupings, and a smaller total RMSE. Overall this model gives a more accurate timing of regime duration in the term structure over the sample period. Lastly, we test to see if both models are statistically different from the non-switching model using a LR test. Our testing results show that the both models are statistically different from the non-switching model at the one percent confidence level.

REFERENCES

Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, 1383-1414.

Breitung, J., & Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, *163*(1), 71-84.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., & Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *The journal of finance*, *47*(3), 1209-1227.

Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International economic review*, 969-996.

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2009). An arbitrage-free generalized Nelson–Siegel term structure model. *The Econometrics Journal*, *12*(3), C33-C64.

--------------------- ------------------- -------------------- . (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics*, *164*(1), 4-20.

Coroneo, L., Nyholm, K., & Vidova-Koleva, R. (2011). How arbitrage-free is the Nelson–Siegel model? *Journal of Empirical Finance*, *18*(3), 393-407.

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

Dai, Q., & Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, *55*(5), 1943-1978.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *64*(2), 247-254.

----------------. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *74*(1), 33-43

Dewachter, H., & Lyrio, M. (2008). Learning, macroeconomic dynamics and the term structure of interest rates. In *Asset Prices and Monetary Policy* (pp. 191-245). University of Chicago Press.

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

Diebold, F. X., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interactions: a dynamic Nelson–Siegel approach. *Journal of Econometrics*,*146*(2), 351-363.

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

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models.*The Journal of Finance*, *57*(1), 405-443.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, *6*(4), 379-406.

Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 680-692.

Garcia, R., & Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 111-125.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357-384.

-----------------. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, *45*(1), 39-70.

Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics*,*7*(S1), S61-S82.

----------------. (1996). Erratum: the likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics*, *11*(2), 195-198.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35-45.

Kim, C.J. (1994). Dynamic Linear Models with Markov-Switching, *Journal of Econometrics*, 60, 1-22.

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

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

Startz, R., & Tsang, K. P. (2010). An Unobserved Components Model of the Yield Curve. *Journal of Money, Credit and Banking*, *42*(8), 1613-1640.

Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident economic indicators. In Leading economic indicators: New approaches and forecasting records, ed. K. Lahiri and G. H. Moore. New York: Cambridge University Press.

----------------   -----------------. (2008). The evolution of national and regional factors in US housing construction. *Princeton University*.

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

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

Wong, E., Lucia, K., Price, S., & Startz, R. (2011). The changing relation between the Canadian and US yield curves. *Journal of International Money and Finance*, *30*(6), 965-981.

Xiang, J., & Zhu, X. (2013). A Regime-Switching Nelson–Siegel Term Structure Model and Interest Rate Forecasts. *Journal of Financial Econometrics*, *11*(3), 522-555.

Yu, W. C., & Salyards, D. M. (2009). Parsimonious modeling and forecasting of corporate yield curve. *Journal of Forecasting*, *28*(1), 73-88.

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

Table 1: Descriptive Statistics

Maturity	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	6.44	2.58	2.69	16.16	0.966	0.692	0.296
3	6.75	2.66	2.73	16.02	0.973	0.714	0.313
6	6.98	2.66	2.89	16.48	0.974	0.730	0.350
9	7.10	2.64	2.98	16.39	0.974	0.738	0.375
12	7.20	2.57	3.11	15.82	0.973	0.742	0.359
15	7.31	2.52	3.29	16.04	0.974	0.750	0.423
18	7.38	2.50	3.48	16.23	0.975	0.757	0.442
21	7.44	2.49	3.64	16.18	0.976	0.761	0.456
24	7.46	2.44	3.78	15.65	0.976	0.758	0.465
30	7.55	2.37	4.04	15.40	0.976	0.768	0.488
36	7.63	2.34	4.20	15.77	0.978	0.774	0.500
48	7.77	2.28	4.31	15.82	0.978	0.780	0.524
60	7.84	2.25	4.35	15.00	0.981	0.794	0.543
72	7.96	2.22	4.38	14.98	0.981	0.803	0.556
84	7.99	2.18	4.35	14.98	0.981	0.786	0.560
96	8.05	2.17	4.43	14.94	0.983	0.812	0.574
108	8.08	2.18	4.43	15.02	0.983	0.813	0.580
120	8.05	2.14	4.44	14.93	0.983	0.789	0.570
Level	7.32	2.31	4.08	15.26	0.980	0.771	0.466
Slope	-1.60	1.45	-5.00	3.33	0.910	0.342	-0.104
Curvature	0.43	0.84	-1.80	3.93	0.736	0.207	0.124

Note: We define the Level as (y(1)+y(24)+ y(120))/3, the Slope as y(120) – y(1) and Curvature as 2*y(24) – (y(1) + y(120))

Table 2: Parameter Estimates

	DNS	DNS-MSL	DNS-MSV
Parameters			
a_{11}	0.95	0.99	0.99
	(0.0178)	(0.0045)	(0.0054)
a_{22}	0.91	0.98	0.99
	(0.0256)	(0.0150)	(0.0094)
a_{33}	0.88	0.79	0.90
	(0.0569)	(0.0322)	(0.0244)
μ_{11}	9.22	7.57	6.85
	(0.6655)	(0.5789)	(0.4108)
μ_{22}	-1.50	-1.52	0.83
	(0.1858)	(0.5368)	(0.3913)
μ_{33}	-0.47	0.16	-0.66
	(0.0543)	(0.2409)	(0.4098)
$\sigma^0_{\eta_L}$	0.36	0.33	**0.50**
	(0.0225)	(0.0146)	(0.0197)
$\sigma^0_{\eta_S}$	0.62	0.61	**1.22**
	(0.0169)	(0.0248)	(0.0212)
$\sigma^0_{\eta_C}$	0.92	1.03	**1.87**
	(0.0283)	(0.0521)	(0.0438)
$\sigma^1_{\eta_L}$	------	------	**0.26**
			(0.0534)
$\sigma^1_{\eta_S}$	------	------	**0.33**
			(0.0686)
$\sigma^1_{\eta_C}$	------	------	**0.61**
			(0.1275)
λ^0	0.080[22.4]	**0.153**[11.7]	0.081[22.1]
	(0.0035)	(0.0022)	(0.0017)
λ^1	------	**0.055**[32.6]	------
		(0.0043)	
q	------	0.90	0.89
p	------	0.93	0.97
Max. Likelihood Value	9243.6	9481.7	9435.3
Number of free parameters (k)	28	31	33
AIC	-18431.2	**-18901.4**	-18804.6
BIC	-18240.6	**-18690.4**	-18578.0

Standard errors are in parentheses
Implied maturities are in brackets
We calculate the information criteria according to the formulas $AIC = -2 * \ell(\boldsymbol{\theta})_{max} + 2 * k$ and $BIC = -2 * \ell(\boldsymbol{\theta})_{max} + k * \ln(NT)$, where $N = 18$ maturities and $T = 371$ months

Table 3: Correlations

	DNS	DNS-MSL	DNS-MSV
ρ(Empirical Level, Estimated. Level)	0.8987	0.8598	0.8768
ρ(Empirical Slope, Estimated Slope)	0.9558	0.9064	0.9061
ρ(Empirical Curvature, Estimated Curvature)	0.8749	0.6374	0.7429

Table 4
Treasury yields in-sample evaluation: Root Mean Squared Errors (RMSEs).

	DNS	DNS-MSL	DNS-MSV
Maturity	RMSE	RMSE	RMSE
1	0.7296	**0.6506**	0.7274
3	0.4919	**0.4274**	0.4840
6	0.2524	0.2875	**0.2302**
9	0.3087	0.3255	**0.3101**
12	0.3253	**0.3219**	0.3272
15	0.3147	**0.2901**	0.3123
18	0.2928	**0.2702**	0.2870
21	0.2777	0.2727	**0.2719**
24	**0.2669**	0.2774	0.2704
30	**0.2702**	0.2723	0.2741
36	0.2831	**0.2712**	0.2890
48	0.3201	**0.3047**	0.3233
60	0.3024	**0.2847**	0.3059
72	0.3243	0.3271	**0.3223**
84	0.3282	0.3350	**0.3278**
96	0.3264	0.3294	**0.3249**
108	0.3879	**0.3756**	0.3898
120	**0.4122**	0.4140	0.4145
Average	0.3453	**0.3354**	0.3440
1-15mo	0.4038	**0.3838**	0.3985
18-48mo	0.2851	**0.2781**	0.2860
60-120mo	0.3469	**0.3443**	0.3475

Table 5
LR Critical Values (1000 Bootstrap Iterations)

	10%	5%	1%	LR-stat
DNS-MSL	22.89	37.58	55.18	467
DNS-MSV	24.64	37.06	55.90	410

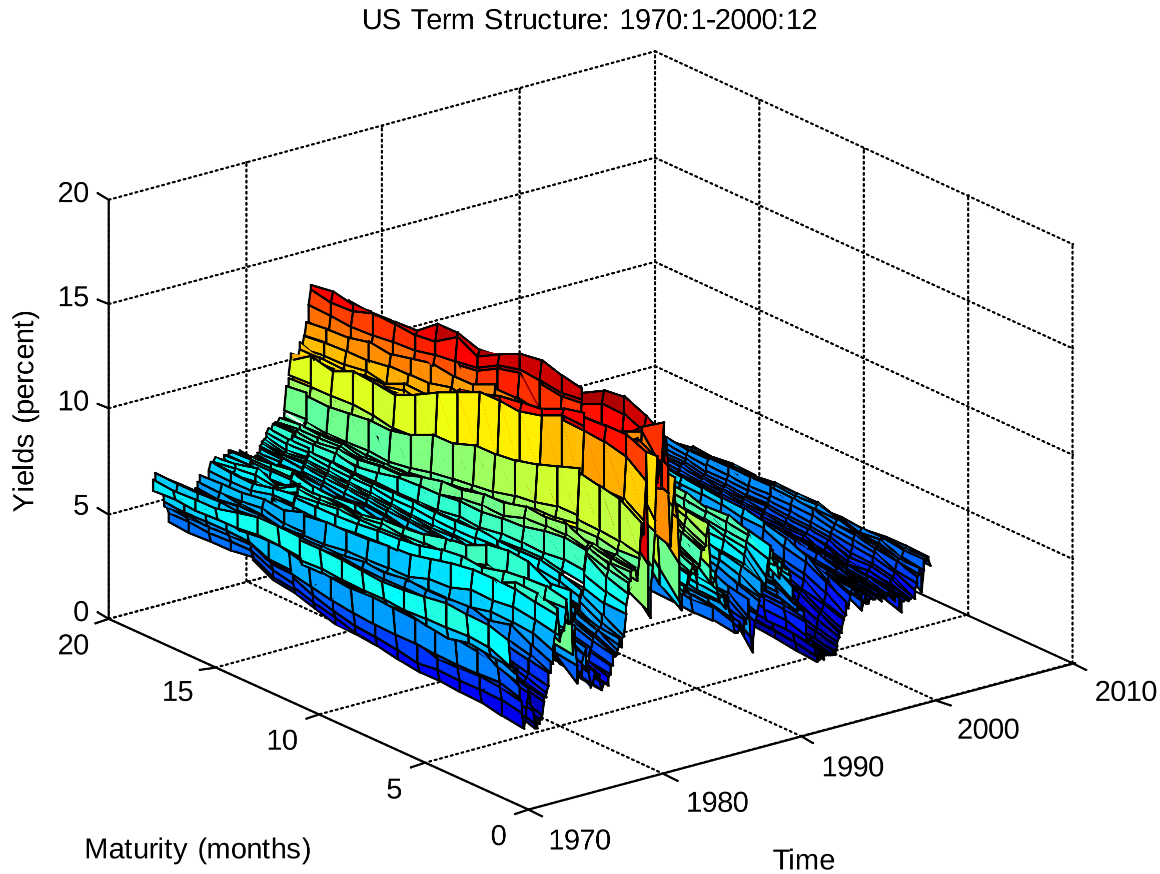

US Term Structure: 1970:1-2000:12
Yields (percent)
20
15
10
5
0
Maturity (months)
20
15
10
5
0
Time
1970
1980
1990
2000
2010

Figure 1

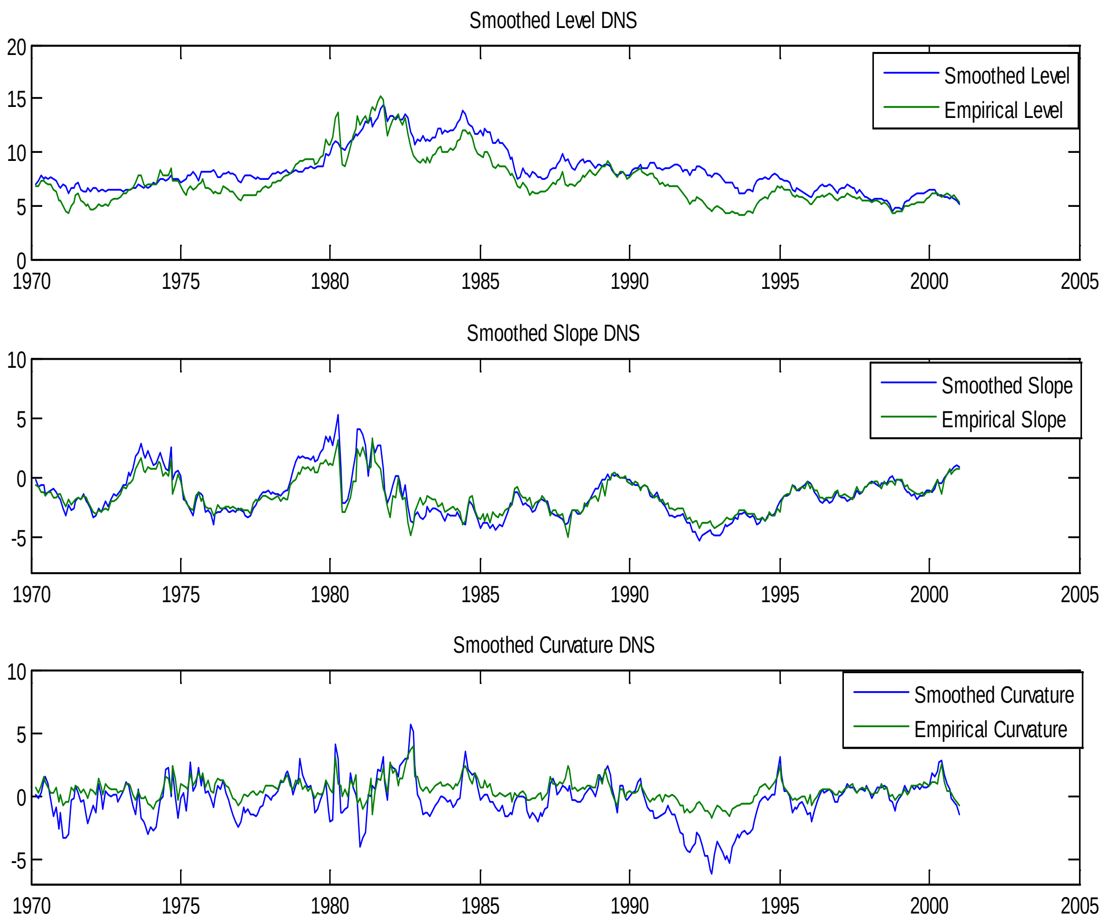

Smoothed Level DNS
Smoothed Level
Empirical Level
20
15
10
5
0
1970
1975
1980
1985
1990
1995
2000
2005
Smoothed Slope DNS
Smoothed Slope
Empirical Slope
10
5
0
-5
Smoothed Curvature DNS
Smoothed Curvature
Empirical Curvature

Figure 2

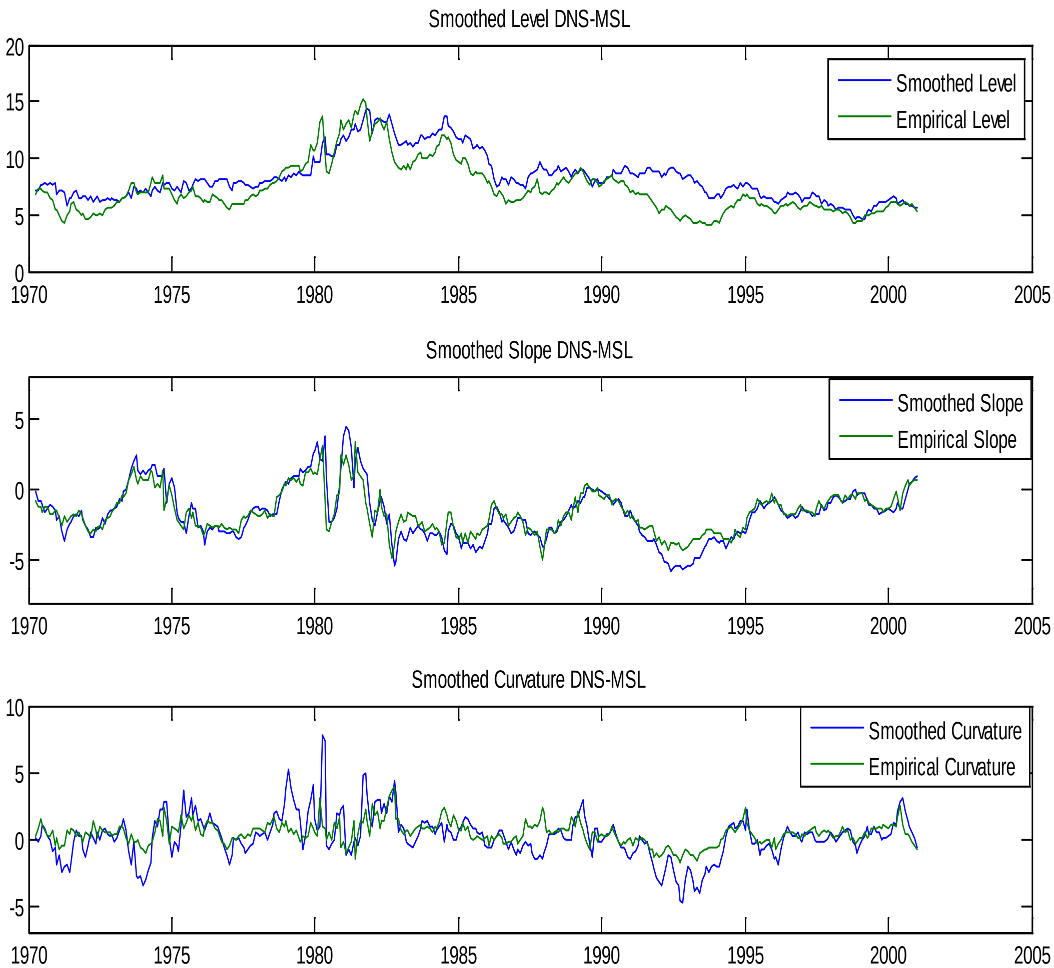

Smoothed Level DNS-MSL
Smoothed Level
Empirical Level
20
15
10
5
0
1970
1975
1980
1985
1990
1995
2000
2005
Smoothed Slope DNS-MSL
Smoothed Slope
Empirical Slope
5
0
-5
1970
1975
1980
1985
1990
1995
2000
2005
Smoothed Curvature DNS-MSL
Smoothed Curvature
Empirical Curvature
10
5
0
-5
1970
1975
1980
1985
1990
1995
2000
2005

Figure 3

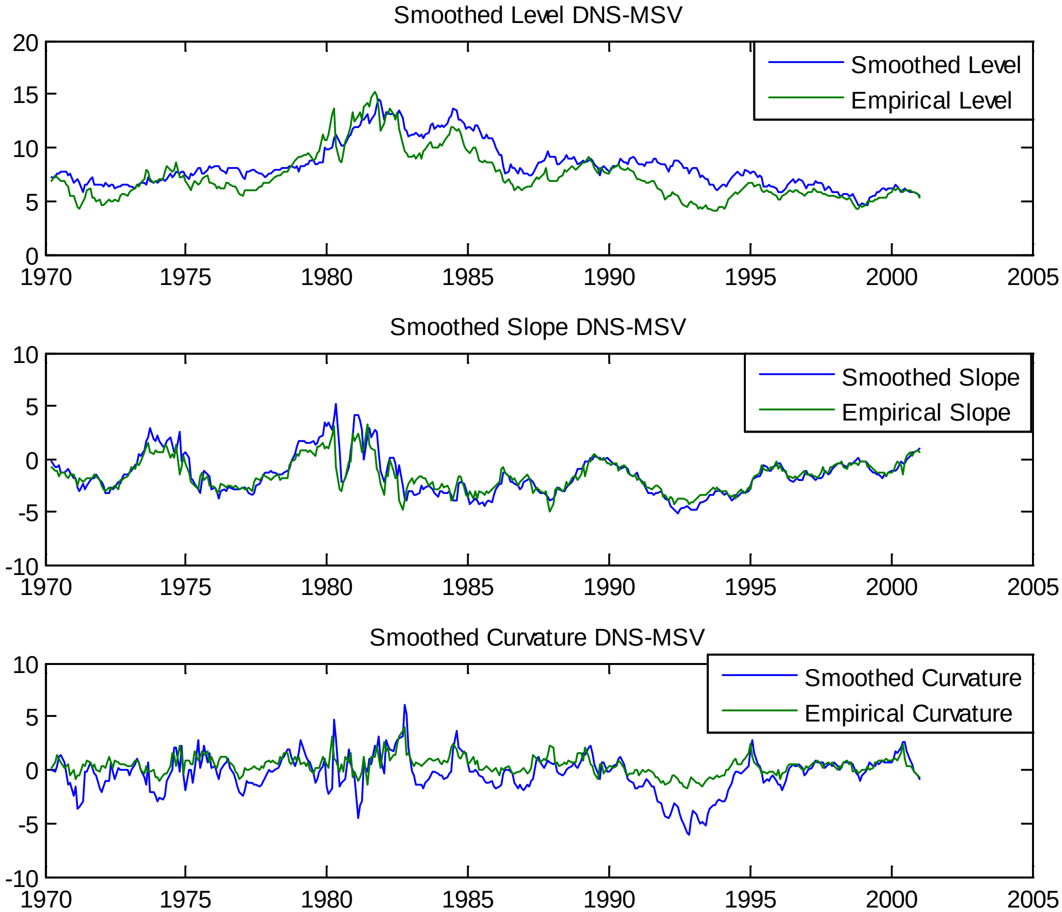

Smoothed Level DNS-MSV
Smoothed Level
Empirical Level
20
15
10
5
0
1970
1975
1980
1985
1990
1995
2000
2005
Smoothed Slope DNS-MSV
Smoothed Slope
Empirical Slope
10
5
0
-5
-10
1970
1975
1980
1985
1990
1995
2000
2005
Smoothed Curvature DNS-MSV
Smoothed Curvature
Empirical Curvature
10
5
0
-5
-10
1970
1975
1980
1985
1990
1995
2000
2005

Figure 4

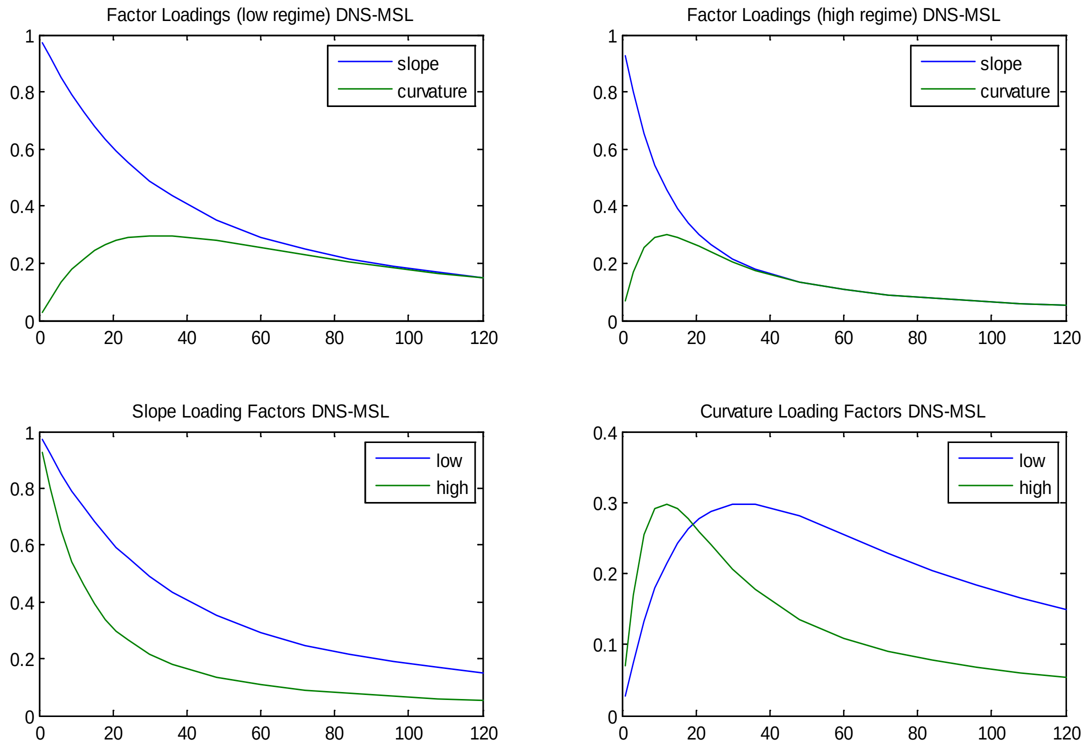

Factor Loadings (low regime) DNS-MSL
slope
curvature
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
Factor Loadings (high regime) DNS-MSL
slope
curvature
Slope Loading Factors DNS-MSL
low
high
Curvature Loading Factors DNS-MSL
low
high
0.4
0.3
0.2
0.1

Figure 5

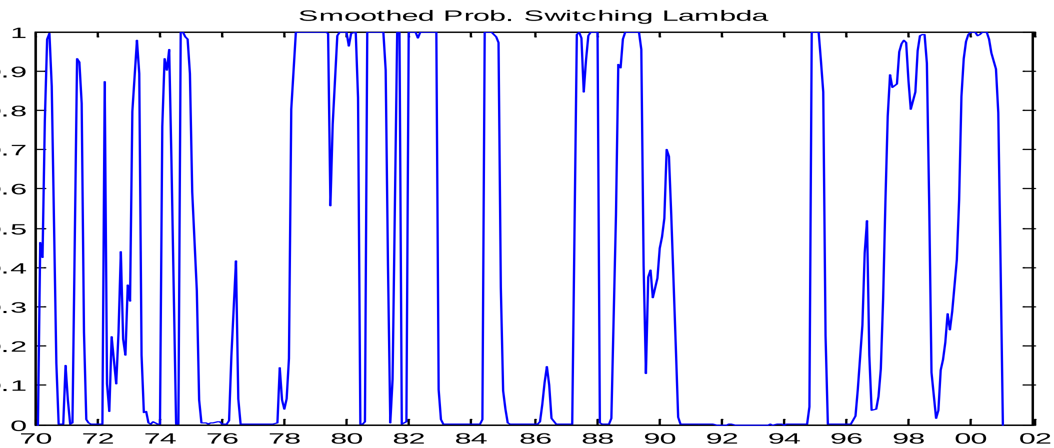
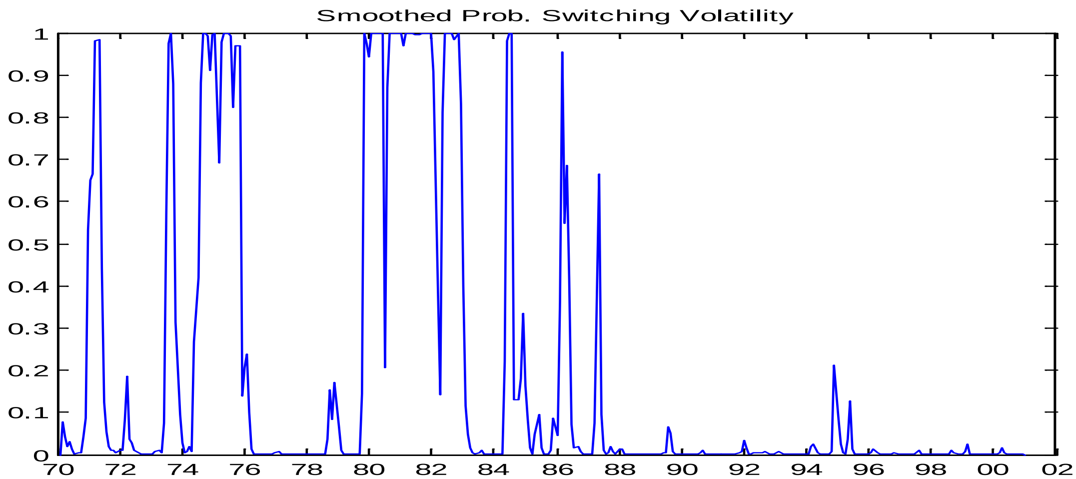
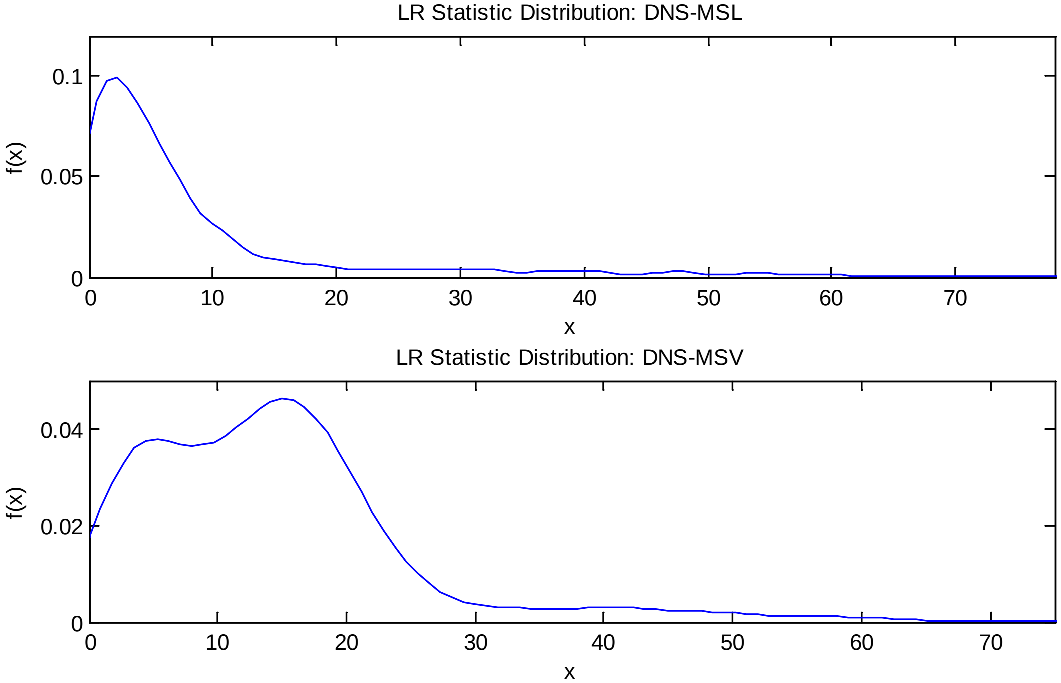

Smoothed Prob. Switching Lambda
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
70
72
74
76
78
80
82
84
86
88
90
92
94
96
98
00
02

Figure 6



Smoothed Prob. Switching Volatility
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
70
72
74
76
78
80
82
84
86
88
90
92
94
96
98
00
02

Figure 7



LR Statistic Distribution: DNS-MSL
f(x)
0.1
0.05
0
0
10
20
30
40
50
60
70
x
LR Statistic Distribution: DNS-MSV
f(x)
0.04
0.02
0
0
10
20
30
40
50
60
70
x

Figure 8