A dynamic Nelson-Siegel yield curve model with Markov switching

Jared Levantᵃ,⁎, Jun Maᵇ

ᵃ Regions Bank, Birmingham, AL, USA
ᵇ Culverhouse College of Commerce & Business Administration, University of Alabama, Tuscaloosa, AL, USA

ARTICLE INFO

JEL:
C51
E43

Keywords:
Nelson–Siegel yield curve model
Regime shifts
State-Space model
Kalman filter
Kim algorithm

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 and that changes in a single parameter that determines the factor loadings of the model better captures regime changes. We show that this model gives superior in-sample forecasting performance as compared to a baseline model and a volatility-switching model. In general, we find compelling evidence that the extracted factors from our term structure models are closely related with various economic variables. Furthermore, we investigate and find evidence that the effects of macroeconomic phenomena such as monetary policy, inflation expectations, and real economic activity differ according to the particular regime realized for the term structure. In particular, we identify the periods where monetary policy appears to have a greater effect on the yield curve, and the periods where inflation expectations seem to have a greater effect in yield determination. We also find convincing evidence of a relationship between the regimes estimated by the various switching models with economic activity and monetary policy.

1. Introduction

The yield curve often contains useful information about real economic activity and inflation. For example, the level factor (the long-term yield-to-maturity) is often argued to be closely related with inflation expectations, 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 is heavily influenced by monetary policy (see Evans and Marshall (1998), and Wu (2002)). The most recent monetary policies, such as 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 interest rate and raise the short-term rate, directly work on the yields by the Federal Reserve Bank in an attempt to lower the long-term yield-to-maturity. The most recent monetary policies, such as 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 interest rate and raise the short-term rate, directly work on the yields by the Federal Reserve Bank in an attempt to lower the long-term.

The interaction of the term structure and the macroeconomy has been investigated by a growing work of empirical literature. Examples include (but are not limited to) Diebold and Li (2006), hereafter DL) and Diebold et al. (2006), hereafter DRA) who employ a generalized version of the Nelson and Siegel (1987), hereafter NS) yield curve model. More recently, Christensen et al. (2011) place the NS model in a theoretically consistent arbitrage-free framework. In this paper, we rely on the results of Coroneo et al. (2011) who finds the NS model is close to being arbitrage-free when applied to the US market, although it does not explicitly impose these restrictions.

Another stream of literature has shown the US interest rate dynamics of the term structure to be subject to frequent regime changes (see Bansal and Zhou (2002)). Although some 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 many other regime changes that are due to more frequent business cycle fluctuations and often indirectly observed changes in the financial markets. Chauvet (1998) introduces regime switching to a dynamic factor model of business cycle fluctuations and thus accurately captures asymmetries associated with economic expansions and contractions. Startz and Tsang (2010) incorporate Markov regime switching into an trend/cycle unobserved components model of the yield curve to account for regime changes of the yield curve. Abdymomunov and Kang (2015) find the differences in...
the term spread across regimes is explained through the term premia rather than expectations of future short rates. Capturing regime changes in order to model the dynamic movements of the yield curve more accurately is becoming a growing source of investigation in the term structure literature as seen in Xiang and Zhu (2013) and Hevia et al. (2015). In particular, our paper differs from Hevia et al. (2015) in a number of important dimensions. First, Hevia et al. (2015) did not report the estimation results of the model with Markov-switching volatility while we find such a model provides important insights for modeling the yields curve. We also conducted statistical exercises to account for the Davies’ nuisance parameter issue in efforts to formally test for the significance of the Markov-switching model relative to the baseline no-switching model (Section 4.5), which Hevia et al. did not attempt to do. Finally, we document the important connections between the regime switchings and macroeconomic indicators using logit models in Section 4.6. In sum, our work made a number of important contributions that are outside the scope of Hevia et al. (2015).

In this paper, we model the parameter instability in the term structure and relate the regime switching to economic fundamentals by applying a Markov-switching component to the factor loading parameter which controls the influence of the slope and curvature yield factors on yields. As mentioned previously, the literature has related the slope factor to monetary policy. Also, it has been shown that the curvature factor is heavily influenced by monetary policy as well (see Dewachter and Lyrio (2008) and Bekar et al. (2010)). In the extant literature, concerning the factor loading parameter, this parameter has been primarily utilized in improving the forecasting ability of the NS model (see Svensson (1995)), Christensen et al. (2009), Koopman et al. (2010)). By assuming the factor loading parameter follows a two-state Markov-process we are able to improve the forecasting ability of the NS model while gaining insight into regime changes of the term structure through the macroeconomic fundamental variables, inflation expectations and monetary policy. Recently, Yu and Salyards (2009) and Yu and Zivot (2011) apply a dynamic NS model to modeling corporate bond yields and they find that the optimal factor loading parameter, changes as one goes from modeling investment to speculative grade bonds. Their results corroborate our findings in general.

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. We apply a Markov switching component to the loading parameters of the factors as well as the factors’ volatility. Comparisons between the models are made by presenting goodness-of-fit statistics and AIC/BIC values. We also implement the Likelihood Ratio (LR) tests to investigate if our models are statistically different from the baseline linear DL model. Although both models are found to be statistically different from the baseline model, the root mean square error (RMSE) analysis shows the model with the loading parameter switching yields the smallest RMSE 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.

In light of recent discussions about potential interactions between the interest rates factors and the macro-economy, we investigate the relationship between the extracted factors from our DNS models and the observed macroeconomic variables. We find that our interest rate factors, which are extracted separately from the macroeconomic variables, are closely related with the macro-economy. Specifically, we find the level factor is strongly correlated with the inflation expectation, and the slope factor appears to be counter-cyclical, which is consistent with the finding by Wu (2002) that the slope factor is related with monetary policy. Furthermore, in the regime-switching DL model we find that the loading of the slope factor on the yield curve is larger during recessions than expansions. This seems to suggest an asymmetric effect of the monetary policy on the yield curve over business cycles.

This paper is organized as follows. Section 2 describes the baseline dynamic Nelson-Siegel model and the regime-switching DNS model. Section 3 describes the data. Section 4 presents and discusses the estimation results. Section 5 concludes. Appendices discuss the estimation procedure via Kalman filter (KF) and the Kim algorithm (KA).

2. Models and estimation

In this section we introduce the baseline dynamic Nelson-Siegel (DNS) model. The appeal of the DNS model lies in its extension to the time dimension. We also introduce our regime-switching extensions of the DNS 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

$$\chi(t) = \gamma_t + \lambda_t \eta_t + \varepsilon_t$$

where $F_t = (L_t, S_t, C_t)$ is a vector representing level, slope, and curvature of the yield curve, for given time $t$, maturity $m$, and constant $\lambda$, 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 $F$. From Eq. (1), the factor loading associated with $L_t$ is assumed to be unity for all maturities and therefore influences short, medium, and long-term interest rates equally. The loading factors for $S_t$ and $C_t$ 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 $\lambda \rightarrow \infty$ (or $m \rightarrow \infty$). The curvature factor loading converges to zero as $\lambda \downarrow 0$ (or $m \downarrow 0$) and as $\lambda \rightarrow \infty$ (or $m \rightarrow \infty$) 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 $\lambda \rightarrow \infty$ 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 both “reasonable” values for $\lambda$ and the level factor are responsible for the wide range of non-flat yield curve shapes within an NS framework.

2.2. DNS model estimation

We adopt the DRA state-space framework to model each variant of the NS model in this paper. Our measurement equation models the yield curve shapes within

$$\begin{align*}
\gamma_t &= \chi_t + \eta_t \\
\eta_t &= \lambda_t \chi_t + \varepsilon_t
\end{align*}$$

or expressed in matrix notation as

$$\begin{pmatrix}
\gamma_t \\
\eta_t
\end{pmatrix} = \begin{pmatrix}
\frac{1}{L_t} & 0 & \lambda_t \\
0 & \frac{1}{S_t} & 0 \\
0 & 0 & \frac{1}{C_t}
\end{pmatrix}
\begin{pmatrix}
\chi_t \\
\lambda_t \\
\varepsilon_t
\end{pmatrix}
$$

with $\chi_t$ representing the $N \times 1$ vector of yields, $N \times 3$ factor loading matrix $A(\lambda), 3 \times 1$ latent factor vector $F_t$, and $N \times 1$ disturbance vector $\varepsilon_t$ (or so-called measurement errors of the yields). The diagonal structure of $\Sigma_t$ implies that measurement errors across maturities of $\chi_t$ 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 process of order one

\[
\begin{pmatrix}
L_{t-1} - \mu_L \\
S_{t-1} - \mu_S \\
C_{t-1} - \mu_C
\end{pmatrix}
= \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}
\begin{pmatrix}
L_t - \mu_L \\
S_t - \mu_S \\
C_t - \mu_C
\end{pmatrix} + \begin{pmatrix} \eta_t \\ \eta_t \\ \eta_t \end{pmatrix}
\]  

(4)

or expressed in matrix notation as

\[
(F_t - (I - A)\mu) = A(F_{t-1} - (I - A)\mu) + \eta_t,
\]

(5)

with 3 x 1 mean vector $\mu$, 3 x 3 coefficient matrix $A$, and 3 x 1 factor disturbance matrix $\Sigma$. In this investigation we assume a diagonal coefficient matrix based on the findings of Diebold et al. (2006) and Christensen, Diebold, and Rudebusch (2011) who show the off-diagonal elements of the $A$ matrix are not statistically relevant to modeling the term structure. We also assume $\Sigma$ follows a diagonal structure implying the factor disturbances are uncorrelated. Once again we look to Diebold et al. (2006), which finds the off-diagonal elements of the factor covariance matrix to be marginally significant. Furthermore, they note that when the model is estimated with the restriction that the factor covariance matrix be diagonal, the point estimates and standard errors of the $A$ matrix are little changed from when the model is estimated with no restriction on $\Sigma$.

Since the DNS state space model is linear in latent factors, we are able to use the Kalman (1960) filter to estimate the latent factors conditional on past and contemporaneous observations of the yields. Appendix A gives a detailed description of the filtering process.

2.3. DNS model with Regime-Switching loading parameter

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

We model $\lambda$ 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 $\lambda$ as a regime switching parameter solely determined by the realized state, $\psi_t$ of the yields. The latent Markov component $\psi_t$ 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 $\lambda(\psi)$ in Eq. (3) is replaced by $\lambda(\psi)\psi$ and the resulting measurement equation is $y_t = \lambda(\psi)\mu + \epsilon_t$. Note that since we are not including $\lambda$ in $F_t$, the observation vector of yields is still linear with respect to our latent factor vector. The modified state-space framework for the dynamic Nelson-Siegel model for estimation with regime-switching the loading parameter is as follows:

\[
y_t = \Lambda(\psi_t)\mu + \epsilon_t, \epsilon_t \sim MN(0, \Sigma_t) = 1, \ldots, T, \quad \psi_t = (I - \mu_0) + \epsilon_t \psi_0, \quad \psi_0 = 0, 1
\]

(6)

These equations written in matrix notation constitute the Dynamic Nelson-Siegel with Markov-Switching Lambda (DNS-MSL) model. Details of the estimation process can be found in Appendix B.

2.4. The DNS Model with Regime-Switching Factor Volatilities

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 volatility to the factor disturbances in the transition equation. The modified state-space framework for the dynamic Nelson-Siegel model for estimation with regime-switching the factor volatilities is as follows:

\[
y_t = \Lambda(\lambda_t)\mu + \epsilon_t, \epsilon_t \sim MN(0, \Sigma_t) = 1, \ldots, T, \quad \psi_t = (I - \mu_0) + \epsilon_t \psi_0, \quad \psi_0 = 0, 1
\]

(7)

These equations constitute the Dynamic Nelson-Siegel with Markov-Switching Volatility (DNS-MSV) model. Details of the estimation process can be found in Appendix B.

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. Fig. 1 gives a 3-dimensional mesh plot of the term structure for the sample period and the maturities.

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 fit.
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 with maturity by 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.

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.\(^1\) 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. Empirical results

In this section we present the model estimation, comparative test results, and explore a macro-factor linkage with the DNS, DNS-MSL, and DNS-MSV models. Parameter estimates of each model via KF and KA are presented and discussed. We then look at in-sample estimation through root mean squared error analysis and information criterion calculations. We test to see if the Markov-switching models are significantly different from the baseline model using a bootstrapped LR test. Finally, we use logit regressions to explore if the estimated regimes are related to macro-factors for inflation expectations, economic activity and monetary policy.

#### 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 \( \lambda \) is 0.080 with a standard

### Table 1


<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.44</td>
<td>2.58</td>
<td>2.69</td>
<td>16.16</td>
</tr>
<tr>
<td>3</td>
<td>6.75</td>
<td>2.66</td>
<td>2.73</td>
<td>16.02</td>
</tr>
<tr>
<td>6</td>
<td>6.98</td>
<td>2.66</td>
<td>2.89</td>
<td>16.48</td>
</tr>
<tr>
<td>9</td>
<td>7.10</td>
<td>2.64</td>
<td>2.98</td>
<td>16.39</td>
</tr>
<tr>
<td>12</td>
<td>7.20</td>
<td>2.57</td>
<td>3.11</td>
<td>15.82</td>
</tr>
<tr>
<td>15</td>
<td>7.31</td>
<td>2.52</td>
<td>3.29</td>
<td>16.04</td>
</tr>
<tr>
<td>18</td>
<td>7.38</td>
<td>2.50</td>
<td>3.48</td>
<td>16.23</td>
</tr>
<tr>
<td>21</td>
<td>7.44</td>
<td>2.49</td>
<td>3.64</td>
<td>16.18</td>
</tr>
<tr>
<td>24</td>
<td>7.46</td>
<td>2.44</td>
<td>3.78</td>
<td>15.65</td>
</tr>
<tr>
<td>30</td>
<td>7.55</td>
<td>2.37</td>
<td>4.04</td>
<td>15.40</td>
</tr>
<tr>
<td>36</td>
<td>7.63</td>
<td>2.34</td>
<td>4.20</td>
<td>15.77</td>
</tr>
<tr>
<td>48</td>
<td>7.77</td>
<td>2.28</td>
<td>4.31</td>
<td>15.82</td>
</tr>
<tr>
<td>60</td>
<td>7.84</td>
<td>2.25</td>
<td>4.35</td>
<td>15.00</td>
</tr>
<tr>
<td>72</td>
<td>7.96</td>
<td>2.22</td>
<td>4.38</td>
<td>14.98</td>
</tr>
<tr>
<td>84</td>
<td>7.99</td>
<td>2.18</td>
<td>4.35</td>
<td>14.98</td>
</tr>
<tr>
<td>96</td>
<td>8.05</td>
<td>2.17</td>
<td>4.43</td>
<td>14.94</td>
</tr>
<tr>
<td>108</td>
<td>8.08</td>
<td>2.18</td>
<td>4.43</td>
<td>15.02</td>
</tr>
<tr>
<td>120</td>
<td>8.05</td>
<td>2.14</td>
<td>4.44</td>
<td>14.93</td>
</tr>
</tbody>
</table>

**Empirical Factors**

<table>
<thead>
<tr>
<th></th>
<th>Level</th>
<th>Slope</th>
<th>Curvature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7.32</td>
<td>-1.60</td>
<td>0.43</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>2.31</td>
<td>1.45</td>
<td>0.84</td>
</tr>
<tr>
<td>Minimum</td>
<td>4.08</td>
<td>-5.00</td>
<td>-1.80</td>
</tr>
<tr>
<td>Maximum</td>
<td>15.26</td>
<td>3.33</td>
<td>4.73</td>
</tr>
</tbody>
</table>

#### Table 2

Parameter Estimates for the DNS, DNS-MSL, and DNS-MSV models.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DNS</th>
<th>DNS-MSL</th>
<th>DNS-MSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{11} )</td>
<td>0.95</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>( \alpha_{21} )</td>
<td>(0.0178)</td>
<td>(0.0045)</td>
<td>(0.0054)</td>
</tr>
<tr>
<td>( \alpha_{31} )</td>
<td>0.91</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>( \lambda_m )</td>
<td>(0.0256)</td>
<td>(0.0150)</td>
<td>(0.0094)</td>
</tr>
<tr>
<td>( \mu_{11} )</td>
<td>0.88</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>( \mu_{22} )</td>
<td>(0.0569)</td>
<td>(0.0222)</td>
<td>(0.0244)</td>
</tr>
<tr>
<td>( \mu_{31} )</td>
<td>9.22</td>
<td>7.57</td>
<td>6.85</td>
</tr>
<tr>
<td>( \rho_{LL} )</td>
<td>0.36</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>( \rho_{SL} )</td>
<td>(0.0225)</td>
<td>(0.0146)</td>
<td>(0.0197)</td>
</tr>
<tr>
<td>( \rho_{SC} )</td>
<td>0.62</td>
<td>0.61</td>
<td>0.61</td>
</tr>
<tr>
<td>( \rho_{CL} )</td>
<td>(0.0169)</td>
<td>(0.0248)</td>
<td>(0.0212)</td>
</tr>
<tr>
<td>( \rho_{LC} )</td>
<td>0.92</td>
<td>1.03</td>
<td>1.03</td>
</tr>
<tr>
<td>( \rho_{SC} )</td>
<td>(0.0283)</td>
<td>(0.0521)</td>
<td>(0.0438)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.080[22.4]</td>
<td>0.153[11.7]</td>
<td>0.081[22.1]</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>(0.0035)</td>
<td>(0.0022)</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>-</td>
<td>0.053[32.6]</td>
<td>-</td>
</tr>
<tr>
<td>( q )</td>
<td>-</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>( p )</td>
<td>-</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td>Max. Likelihood Value</td>
<td>9243.6</td>
<td>9481.7</td>
<td>9435.3</td>
</tr>
<tr>
<td>Number of free parameters (k)</td>
<td>28</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>AIC</td>
<td>-18431.2</td>
<td>-18901.4</td>
<td>-18804.6</td>
</tr>
<tr>
<td>BIC</td>
<td>-18240.6</td>
<td>-18690.4</td>
<td>-18578.0</td>
</tr>
</tbody>
</table>

Standard errors are in parentheses.

\( CL = \frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m} \)
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 estimated DNS model is 22.4 months while the implied maturity in the DRA paper is 23.3 months.

Our estimated smoothed factors from the DNS model have relatively high correlations to the empirical factors as shown in Table 3. In fact, the smoothed factors of this model give the highest correlations for each empirical factor of all the models estimated. Fig. 2 gives a visual of the time-series of each estimated factor plot with the time-series of the empirical factors.

### 4.2. DNS-MSL model

We now 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 from Table 3 we find a drop in the correlations across the factors as compared with the baseline non-switching model. Specifically, the curvature factor experiences the largest decrease of all the models. This is further supported with a visual inspection of the third plot of the smoothed-empirical factor plots of Fig. 3.

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 KF indicate the loading parameter \( \lambda \) is subject to a hidden Markov switching component. We estimate \( \lambda \) to be 0.055 (32.6 months) and 0.153 (11.7 months) for the low and high regimes, respectively. Fig. 4 shows the effect these values have on the slope and curvature factor loadings across maturities.

In the first two plots we see a much faster decay of the slope and curvature loading factors in the high \( \lambda \) regime than in the low \( \lambda \) regime. This indicates that in the high \( \lambda \) regime the slope and curvature factor loadings (which are directly dependent on \( \lambda \)) influence medium and long-term maturities less than in the low \( \lambda \) regime. Yields for medium and long-term maturities are determined more by long-term inflation expectations during regimes when \( \lambda \) is relatively high—because the decay of the slope and curvature factors is greater—and more by economic activity and monetary policy when \( \lambda \) is low—because the decay of the slope and curvature factors is relatively less.

From the third plot in Fig. 4, we see that the slope loading factor is uniformly greater across maturities in the low regime than in the high regime. This shows that the slope factor—proxying for monetary policy—contributes more to yield determination over all maturities in the low \( \lambda \) regime relative to inflation expectations. 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

### Table 3

Correlations between the empirical and estimated factors.

<table>
<thead>
<tr>
<th></th>
<th>DNS</th>
<th>DNS-MSL</th>
<th>DNS-MSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_{LEMP,LEST} )</td>
<td>0.8987</td>
<td>0.8598</td>
<td>0.8768</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>( \rho_{SEMP,SEST} )</td>
<td>0.9508</td>
<td>0.9064</td>
<td>0.9061</td>
</tr>
<tr>
<td>(0.004)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>( \rho_{CEMP,CEST} )</td>
<td>0.8749</td>
<td>0.6374</td>
<td>0.7429</td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.028)</td>
<td>(0.023)</td>
<td></td>
</tr>
</tbody>
</table>

Note: \( \rho_{x,y} \) is the correlation coefficient between quantities \( x \) and \( y \). \( LEMP \) is the empirical level factor, \( LEST \) is the estimated level factor, \( SEMP \) is the empirical slope factor, \( SEST \) is the estimated slope factor, \( CEMP \) is the empirical curvature factor, and \( CEST \) is the estimated curvature factor. Standard errors are in parentheses.
factor decays quickly and is less of a factor in yield determination than in the low $\lambda$ regime.

Fig. 5 plots the time series of the of the unobserved state, $\psi_t$, of the DNS-MSL model.

The timing and duration of the regime changes in Fig. 5 coincide very nicely with periods of changing inflation expectations, economic activity and monetary policy. From the figure we see in the 1970s and early 1980s, the high $\lambda$ regime was persistent. Based on our above discussions, these regimes correspond to periods in which inflation expectations weighed most heavily on Treasury yield determination.
After the monetary policy experiment of the late 70s and early 80s, we witness more persistent periods where monetary policy plays a more important role in the bonds market which coincides with the Great Moderation. The peaks associated with a high \( \lambda \) regime are less frequent and less persistent during this period from the late 1980s throughout the 1990s. Gurkaynak and Wright (2012) interpret the 1990s as a period where inflation uncertainty was diminishing. In Fig. 5 we observe that the years from mid-1990 until mid-1994 are clearly categorized as a low \( \lambda \) regime where monetary policy had a stronger effect on the bonds market. This period corresponds to the mild recession caused by the oil supply shock as a result of the Gulf War and the Federal Reserve’s lowering of the federal funds rate to stimulate economic activity.

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 various 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. The estimated factors display a decrease in their correlations with the empirical factors as compared to the correlations between the baseline model’s factors and empirical factors. But the decrease is not as great as that of the DNS-MSL correlations. And once again, the curvature factor experiences the smallest correlation with its empirical counterpart. The time-series of the factors is given in Fig. 6.

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 \( \lambda \) to be 0.081. Our \( \lambda \) estimate gives an estimate to be 0.081. Our regime where monetary policy had a small effect on the bonds market began targeting money aggregates in late 1979 and then switched back to targeting the Federal Funds Rate in late 1982. These changes in monetary policy manifested itself as increased uncertainty in US financial markets during this period. Factor volatilities began to show a sustained decrease in the mid-1980s with ephemeral spikes occurring until 1988 where afterwards the volatilities are in the low regime until the end of the sample period. The timing of this decrease corresponds to the literature’s exploration of the Great Moderation. Kim and Nelson (1999), Stock and Watson (2003), and Fogli and Perri (2006) all find evidence of a decrease in the volatility of US economic activity around 1984, which is close to when the factor volatilities began to show a transition to the long-term low regime.

In accordance with the literature, we estimate a general model where both the decay parameter and factor disturbances are subject to switching, simultaneously. The RMSE results show this model to be superior to the DNS-MSL in-sample fit results but inferior to the DNS-MSL results. We do not include the results of this most general model in this paper but are available upon request.

4.4. In-sample estimation

Table 4 reports the in-sample root mean squared error (RMSE) values for the various models.

Recall that the KF estimates measurement error parameters for each maturity. 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

\[
RMSE = \left( \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 / T \right)^{1/2}
\]

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 DNS-MSV models 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-term 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. 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 uncertainty.
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.

4.5. Statistical difference between the Baseline Model and the switching models

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 of the LR test nonstandard. This issue has been addressed by Davies (1977, 1987), Andrews and Ploberger (1994), and Hansen (1992, 1996) among others. We follow Hansen’s
We outline the steps used to bootstrap the LR distribution for the DNS and DNS-MSL models:

**STEP 1:** Obtain the max likelihood value \( L_{LV_0} \) of the DNS model (null) and the max likelihood value \( L_{LV_A} \) of the DNS-MSL model (alternative), using the real dataset. Calculate the LR statistic.

**STEP 2:** Generate yields \( l_i \) using the DNS model. Fit the DNS-MSL and DNS models to the generated yields \( l_i \). Obtain the max likelihood value \( L_{LV_0}^* \) for the DNS-MSL model and \( L_{LV_A}^* \) for the DNS model. Calculate the new LN*R using \( L_{LV_0}^* \) and \( L_{LV_A}^* \).

**STEP 3:** Using \( L_R \) and \( L_{LN*R}^* \) compute a bootstrap critical value (\( \hat{C}_a^* \)). For a test at level \( a \), first sort the \( L_{LN*R}^* \) from smallest to largest. Then calculate

\[
\hat{C}_a^* \approx L_{LN*R}^{*\min(B+1)}
\]

where \( a \) represents your confidence level and \( B \) is the number of bootstraps. Repeat steps 2 and 3 \( B \) times and obtain \( L_{LN*R}^{*\{B\}} \).

**STEP 4:** Reject the null hypothesis if \( L_R > \hat{C}_a^* \).

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. Fig. 8 is a plot of the probability density for both models using a normal kernel function to smooth. Table 5 lists the critical values for each model at the 10%, 5% and 1% confidence levels.

It is evident that the test statistic greatly exceeds all bootstrapped critical values so we are able to reject the null of the linear DNS model. 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.

4.6. The relationships of the factors to the macro-economy

A number of papers have tried to establish relationships between yield curve factors, such as the NS factors or the first three principal components of the term structure, with macro-factors. We start with a look at correlations between the level and slope factors of our various models with macro-factors in inflation expectations and capacity utilization. Inflation expectations data comes from the University of Michigan survey in the St. Louis FRED database and covers the period January 1978 through December 2000. Capacity utilization which serves as our proxy for economic activity also comes from the FRED database and covers the period January 1970 through December 2000. Table 6 shows the correlations for the smoothed factors of the DNS, DNS-MSL, and DNS-MSV models.

---

**Table 4**

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

Note: The bold-faced RMSEs represent the minimal RMSE for each maturity and maturity ranges.
Correlations between inflation expectations and the smoothed level factors of the various models are about 0.4. This is a good indicator that the level factor is capturing the dynamics of inflation from the late 1970s through 2000. The DNS model yields the highest correlation (0.4135) while the DNS-MSL model gives the smallest (0.3982).

Correlations between smoothed slope factors and capacity utilization are about -0.2. The negative correlation here indicates that recessions are related with a larger magnitude of the slope factor and hence steeper curve. This tells us that the slope factor is more related with counter-cyclical monetary policy. In recessions the Fed tends to lower the short rate (recall the Taylor rule) and thus steepens the yield curve. Although flatter yield curve tends to predict recessions, steeper curve is contemporaneously related with recessions. The DNS model gives the largest absolute value for the correlation (0.1995) while the DNS-MSL gives the smallest absolute value (0.1935).

The investigations of Bansal and Zhou (2002), Clarida et al. (2006), Xiang and Zhu (2013) show a relationship between interest rate regimes and real economic activity. To investigate a similar relationship with our regime switching models, we estimate a logit model using the monthly capacity utilization data as our measure of economic activity and transform smoothed probabilities from the DNS-MSL and DNS-MSV models into a binary probability variable, $\rho(y)$. The binary variable is defined to assume the value of zero when the smoothed probability value is greater than or equal to one-half and one when less than one-half. Thus, we focus on the low regimes. Fig. 9 gives the DNS-MSL smoothed probability plot superimposed on a plot of the inflation expectations and capacity utilization macro-factors for the period 1895–2000.

Fig. 10 gives the DNS-MSV smoothed probability plot superimposed on a plot of the same macro-factors and time period found in Fig. 9.

Fig. 11 is a transformed binary probability plots for the DNS-MSL model superimposed on the inflation expectations and capacity utilization macro-factors.

Fig. 12 gives the DNS-MSV transformed binary probability plots superimposed on a plot of the same macro-factors and time period found in Fig. 11.

The logit model takes the following form

$$p(y_{lw}) = \frac{\exp(\beta_0 + \beta_1 \text{CAPUTIL})}{1 + \exp(\beta_0 + \beta_1 \text{CAPUTIL})}$$

where $p(y_{lw})$ is the transformed smoothed probability of being in the low interest rate level regime for the DNS-MSL model and the low volatility regime for the DNS-MSV model.

The sample period we investigate with the logit regressions is January 1985 through December 2000 to mitigate the interest rate volatility over the period associated with the monetary policy experiment of the late 1970s and early 1980s and to take advantage of the economic stability associated with the Great Moderation.

In times of low interest rate volatility, manufacturers will increase planned investment spending due to the low level of economic uncertainty. Thus we would expect increased economic activity. The DNS-MSV logit model should reflect this relationship with positive coefficients for capacity utilization when regressing on transformed low volatility regime smoothed probabilities. The DNS-MSV logit model gives $\beta_1 = 1.33(0.59)$ which implies the odds of being in the low volatility...
regime increases by 74%. The Mcfadden pseudo-$R^2$ of this regression was 0.31. Xiang and Zhu (2013) also find a significantly positive coefficient for their logit model when the low volatility regime probabilities are used. This result supports the economic prior that being in a low volatility regime is more likely to occur during an economic expansion. Abdymomunov and Kang (2015) further echo this in finding a decrease in yield volatilities when the monetary authority implements "passive" monetary policy during expansions.

The DNS-MSL logit parameter estimates yields interesting results that suggest an asymmetric impact of monetary policy on the yield curve over the business cycles. The logit regression yields a parameter estimate of $\hat{\beta} = -0.58(0.11)$ with a pseudo-$R^2$ of 0.15. This estimate translates to a 79% decrease in the odds of being in the low lambda regime. In other words, increasing capacity utilization or an economic boom tends to be associated with the high $\lambda$ regime, in which the slope factor has a relatively smaller impact on the yields curve compared to the low $\lambda$ regime (recall the bottom left plot of Fig. 4). Since the slope factor normally approximates the monetary policy this finding suggests that the monetary policy seems to have a larger impact on the yield curve during recessions than expansions. Once again support for this can be found in Abdymomunov and Kang (2015) who show an increase in the term spread due to "active" monetary policy to combat inflationary pressures on economic growth or economic slowdowns. This reveals an interesting asymmetric effect of the monetary policy on

---

3 For the logit regression a sufficient condition for a satisfactorily good model fit is when the McFadden pseudo-$R^2$ falls within the interval [0.2, 0.4].
the economy through the yields curve.

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 the factors’ conditional volatilities show significant switching when allowed—not the conditional mean as noted in the literature. Specifically, the model allowing switching loading parameters 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. We also test to see if both models are statistically different from the non-switching model using a LR test. To correct for non-standard errors due to a nuisance parameter issue, we bootstrap critical values for the test. Our testing results show that both models are statistically different from the non-switching model at the one percent confidence level, thus supporting an extension of the DNS model to include regime switching components.

Finally, we find that the extracted factors from our DNS models are closely related with the macro-economy. In particular, the level factor appears to be strongly correlated with inflation expectations while the slope factor seems to be counter-cyclical, which is consistent with some previous findings, such as Wu (2002), that the slope factor may well be closely related with monetary policy. In addition, we also find that the regime switching in the extended DNS models coincides with economic activity and monetary policy changes. The model accounting for regime changes in volatility captured the timing of volatility regimes associated with the oil price shock of the 1970s, the monetary policy experiment of the early 1980s and the period known as the Great Moderation. The model accounting for regime changes in the loading parameter suggests an interesting asymmetric effect of monetary policy on the yield curve. Specifically, we find that the monetary policy tends to have a larger impact on the yield curve during recessions than expansions.

Appendix A. Kalman Filter

Since the DNS model is linear in latent factors, we are able to use the Kalman filter (KF) to estimate the latent factors conditional on past and contemporaneous observations of the yields. The KF procedure is carried out recursively for \( t = 1, \ldots, T \) with initial values for the latent factors and their variances being the unconditional mean and unconditional variance, respectively. If we define \( f_{t0} \) as the minimum mean square linear estimator (MMSLE) of \( F_t \) and \( v_{t0} \) as the mean square error (MSE) matrix, then \( f_{t0} = \mu \) and \( v_{t0} = (I - A)^\top \Sigma_f \). With observation \( y_t \) and initial values \( f_{10} \) and \( v_{10} \) available, the KF updates the values for \( f_t \) and \( v_t \) using the equations

\[
\begin{align*}
    f_t &= f_{t-1} + K_te_{t-1}, \\
    v_t &= v_{t-1} - K_t \Lambda(\delta)v_{t-1},
\end{align*}
\]  

(A.1)

\[ v_t = v_{t-1} - K_t \Lambda(\delta)v_{t-1}, \]  

(A.2)

where \( e_{t-1} = y_t - \Lambda(\delta)f_{t-1} \) is the predicted error vector, \( v_{t-1} = \Lambda(\delta)v_{t-1} + \Sigma_f \) is the predicted error variance matrix and \( K_t = v_{t-1} \Lambda(\delta)e_{t-1}^{-1} \) is the Kalman gain matrix.

The next period \( t + 1 \) MMSLE of the latent factors and associated variance matrix conditional on yields \( y_{t}, \ldots, y_{T} \) are governed by the prediction equations

\[
\begin{align*}
    f_{t+1} &= (I - A)f_t + Af_{t-1}, \\
    v_{t+1} &= Av_{t-1}A^\top + \Sigma_v.
\end{align*}
\]  

(A.3)

(A.4)
Denote $\theta$ as the system parameter vector. The parameters to be estimated via numerical maximum likelihood estimation are $\theta_{DNS} = \{A, \Sigma, \Sigma_0, \mu, \lambda\}$. We represent the likelihood function as

$$\ell(\theta) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log |w_t| - \frac{1}{2} \sum_{j=1}^{T} \epsilon_j^2 (\epsilon_j)^2.$$  

(A.5)

The function $\ell(\theta)$ 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 $f_0$ and $v_0$ 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:

$$f_{t+1} = f_t + v_{t+1} A^\dagger (f_{t+1} - \lambda(f_t - \mu)),$$

(A.6)

$$v_{t+1} = v_t + A^\dagger (f_{t+1} - \lambda(f_t - \mu) - v_{t+1} A^\dagger (f_{t+1} - \lambda(f_t - \mu))) v_t.$$  

(A.7)

These smoothed estimates provide a more accurate inference on $f_t$ because it uses more information from the system than the filtered estimates.

**Appendix B. Kim Algorithm**

The Kim algorithm (KA) allows for efficient estimation of parameters through the KF and accurate inference of the realized states through a methodology developed by Hamilton (1989, 1990). For both the DNS-MSL and DNS-MSV models, the transition probabilities between states are governed by the entries of the matrix

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

where $p_j = \Pr[\psi_j = 1 | \Omega]$ with $\sum_{j=0}^{1} p_j = 1$ for all $i$.

The estimation of the model parameters according to the KA 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, $\psi_i$. We initialize the NS factors and their variances as in the non-switching case. To initialize the unobserved state, $\psi_i$, we need $\Pr[\psi_i = j | \Omega_i]$ where $j = 0, 1$ and $\Omega_i$ 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

$$
\begin{align*}
\sigma_0 &= \Pr[\psi_0 = 0 | \Omega_0] = \frac{1-p}{2-p-q} \\
\sigma_1 &= \Pr[\psi_0 = 1 | \Omega_0] = \frac{1-q}{2-p-q},
\end{align*}
$$

(B.1)

(B.2)

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

$$f_{i,t}^{(j)} = f_{i,t-1}^{(j)} + K_{i,t-1}^{(j)} e_{i,t-1},$$

(B.3)

$$\psi_{i,t}^{(j)} = \psi_{i,t-1}^{(j)} - K_{i,t-1}^{(j)} A^\dagger (f_{i,t-1}^{(j)}),$$

(B.4)

$$e_{i,t}^{(j)} = \psi_{i,t}^{(j)} - \Lambda(f_{i,t}^{(j)}),$$

(B.5)

$$e_{i,t}^{(j)} = \Lambda(\psi_{i,t}^{(j)} - \Lambda(f_{i,t}^{(j)}) - \Sigma_0,$$

(B.6)

$$f_{i,t-1}^{(j)} = (I - A) f_{i,t-1}^{(j)} + A f_{i,t-1}^{(j)},$$

(B.7)

$$A^{\dagger} f_{i,t-1}^{(j)} = A^{\dagger} f_{i,t-1}^{(j)} + A^{\dagger} f_{i,t-1}^{(j)} + \Sigma_q,$$

(B.8)

where $K_{i,t-1}^{(j)} = w_{i,t-1} A^\dagger (\psi_{i,t-1}^{(j)} - \Lambda(f_{i,t-1}^{(j)}))^{-1}$ is the Kalman gain.

The efficiency of the KA arises from collapsing the $(2 \times 2)$ posteriors $f_{i,t}^{(j)}$ and $\psi_{i,t}^{(j)}$ into two single-state posteriors

$$f_{i,t} = \sum_{j=0}^{1} \Pr[\psi_t = j | \Omega_i] f_{i,t}^{(j)}$$

(B.9)

and

$$\psi_{i,t} = \sum_{j=0}^{1} \Pr[\psi_t = j | \Omega_i] \psi_{i,t}^{(j)}$$

(B.10)

by taking weighted averages over states at $t - 1$. Following Hamilton (1989, 1990), the Kim (1994) algorithm 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

$$\Pr[\psi_t = j, \psi_{t-1} = i | \Omega_i] = \frac{f(\psi_t = j, \psi_{t-1} = i | \Omega_i) \Pr[\psi_t = j | \psi_{t-1} = i | \Omega_{i-1}]}{Pr[\psi_t = j | \Omega_i]},$$

(B.11)

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 $f(\psi_t | \psi_{t-1} = i, \Omega_{t-1})$ is obtained based on the prediction error decomposition:

$$f(\psi_t | \psi_{t-1} = i, \Omega_{t-1}) = (2\pi)^{-\frac{N}{2}} |\psi_{i,t}^{(j)}|^{-\frac{1}{2}} e^{\frac{1}{2} (\psi_{i,t}^{(j)})^T \psi_{i,t}^{(j)}}$$

$$\exp\left\{ \frac{1}{2} e_{i,t}^{(j)} (\psi_{i,t}^{(j)} - \Lambda(f_{i,t}^{(j)})^{-1} e_{i,t}^{(j)} \right\}.$$
and
\[
\Pr \{ \psi_j = i, \psi_{\ell} = j | \Omega_{-1} \} = \Pr \{ \psi_j = i \} \times \Pr \{ \psi_{\ell} = j | \Omega_{-1} \}
\]
where \( \Pr \{ \psi_j = i, \psi_{\ell} = j \} \) is the transition probability. The terms in the numerator are now in known terms. The denominator, \( \Pr \{ \psi_j | \Omega_{-1} \} \), can be expressed as
\[
\Pr \{ \psi_j = i, \psi_{\ell} = j | \Omega_{-1} \} = \sum_{i=0}^{T} \sum_{\ell=0}^{T-1} \Pr \{ \psi', \psi_j = i, \psi_{\ell} = j | \Omega_{-1} \}.
\]
(B.12)

Finally, summing over state \( \ell \) we get our single state posterior
\[
\Pr \{ \psi_j = i | \Omega_{-1} \} = \sum_{\ell=0}^{T} \Pr \{ \psi_j = i, \psi_{\ell} = j | \Omega_{-1} \}.
\]
(B.13)

From the filter we obtain the density of \( \psi \) conditional on past information \( \psi_{\ell} = i, \ell = 1, 2, \ldots, T \). We can now calculate maximum likelihood estimates from the approximate log likelihood function
\[
\ell(\theta) = \ln \{ \Pr \{ \psi_j = i, \psi_{\ell} = j | \Omega_{-1} \} \} = \sum_{i=0}^{T} \sum_{\ell=0}^{T-1} \ln \left( \Pr \{ \psi_j = i, \psi_{\ell} = j | \Omega_{-1} \} \right).
\]
(B.14)

Because these are switching models, the parameter vector set for both are going to be have more parameters estimated than the non-switching model: \( \theta_{DNS-MV} = \{ A_{ij}, \Sigma_{ij}, \mu, \lambda_{ij} \} \) and \( \theta_{DNS-MSV} = \{ A_{ij}, \Sigma_{ij}, \mu, \lambda_{ij} \} \).

Once we have finished calculating the maximum of \( \ell(\theta) \), all parameters have been estimated and we can get inferences on \( \psi_j \) and \( \psi_{\ell} \) conditional on all the information in the sample: \( \Pr \{ \psi_j = i | \Omega_{-1} \} \) and \( \psi_{\ell} = j, \ell = 1, 2, \ldots, T \). 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
\[
\Pr \{ \psi_j = i, \psi_{\ell} = j | \Omega_{-1} \} \approx \Pr \{ \psi_j = i | \Omega_{-1} \} \times \Pr \{ \psi_{\ell} = j | \Omega_{-1} \}
\]
(B.15)

and probability
\[
\Pr \{ \psi_j = i | \Omega_{-1} \} = \sum_{k=0}^{T} \Pr \{ \psi_j = i, \psi_{\ell} = k | \Omega_{-1} \}
\]
(B.16)

These probabilities are used as weights in weighted averages to collapse the \((M \times M)\) elements of \( f_{ij}^{t,k} \) and \( \psi_{\ell}^{t,k} \) into \( M = 2 \) for our model. These weighted averages over \( \psi_{\ell}^{t,k} \) are
\[
f_{ij}^{t,k} = \sum_{k=0}^{T} \Pr \{ \psi_j = i | \Omega_{-1} \} \times \Pr \{ \psi_{\ell} = j | \Omega_{-1} \} \times \Pr \{ \psi_j = i, \psi_{\ell} = k | \Omega_{-1} \}
\]
(B.17)

and
\[
\psi_{\ell}^{t,k} = \sum_{i=0}^{M} \Pr \{ \psi_j = i | \Omega_{-1} \} \times \Pr \{ \psi_{\ell} = k | \Omega_{-1} \} \times \Pr \{ \psi_j = i, \psi_{\ell} = k | \Omega_{-1} \}
\]
(B.18)

Taking a weighted average over the states at time \( t \) we get an expression for the smoothed factors
\[
f_{ij}^{t,k} = \sum_{\ell=0}^{T} \Pr \{ \psi_j = i | \Omega_{-1} \} \times \Pr \{ \psi_{\ell} = j | \Omega_{-1} \}
\]
(B.19)

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

References


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

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74 (1), 33–43.


J. Levant, J. Ma
